5-2012

Environmental Assessment of Streams: Linking Land Use, Instream Stressors, and Biological Indices to Infer Likely Causes of Ecological Impairment

Jacob J. Vander Laan
Utah State University

ENVIRONMENTAL ASSESSMENT OF STREAMS: LINKING LAND USE, INSTREAM STRESSORS, AND BIOLOGICAL INDICES TO INFER LIKELY CAUSES OF ECOLOGICAL IMPAIRMENT

by

Jacob J. Vander Laan

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Ecology

Approved:

Charles Hawkins Major Professor	Patrick Belmont Committee Member
Richard Cutler Committee Member	Mark R. McClellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2012

ABSTRACT

Environmental Assessment of Streams: Linking Land Use, Instream Stressors, and Biological Indices to Infer Likely Causes of Ecological Impairment

by

Jacob J. Vander Laan, Master of Science

Utah State University, 2012

Major Projessor: Dr. Charles P. Hawkins
Department: Watershed Sciences

To protect and restore the biological integrity of streams, we need to be able to both detect biological degradation and infer likely causes of impairment. Managers often use biological indices to measure biological condition and detect degradation. However, the ability to detect degradation can be limited by the performance of the indices we develop. Index performance varies widely, but the sources of this variation are often unclear. In addition, although bioassessments are useful tools for detecting biological degradation, they do not identify stressors associated with impairment. My thesis research had two general goals: 1) develop statistically and ecologically robust indices to measure biological condition in Nevada streams and 2) quantify relationships between land uses, stressors, and biological condition to infer likely causes of degradation.

I developed two biological indices for Nevada streams, a multimetric index (MMI) and observed to expected (O/E) taxa ratios, and determined if index performance was related to site isolation and sample evenness. The Nevada O/E indices were relatively imprecise compared

with those from other regions, which likely results from low assemblage predictability associated with spatial isolation of aquatic habitats in arid regions. In contrast, the Nevada MMI was more precise than most previously developed MMIs, likely the result of using models to reduce natural variation in index scores. Sample evenness was positively associated with both O/E and MMI scores. Adjustments of index scores for sample evenness increased index precision, but also altered relative differences in index values and therefore inferences of biological impairment at specific sites.

I also quantified relationships between biological condition, instream stressors, and land uses and used a weight of evidence approach to infer likely causes of degradation. Land uses such as agriculture, urbanization, and mining were associated with the spatial distributions of instream stressors, and these stressors were associated with variation in biological condition. Total dissolved solids and metal contamination were the stressors most strongly associated with biological condition. By detecting biological degradation and identifying important stressors and their potential sources, the tools I developed should help managers target conservation and restoration efforts and improve their ability to protect freshwater resources.

(100 pages)

PUBLIC ABSTRACT

Environmental Assessment of Streams: Linking Land Use, Instream Stressors, and Biological Indices to Infer Likely Causes of Ecological Impairment

by

Jacob J. Vander Laan, Master of Science

Utah State University, 2012

Major Projessor: Dr. Charles P. Hawkins
Department: Watershed Sciences

To protect and restore stream biota, managers need to be able to both detect biological degradation and infer likely causes of impairment. Managers commonly develop indices based on stream biota to assess the biological condition of streams. However, the performance of these indices dictates the ability to detect degradation, and although index performance varies widely, the sources of this variation are often unclear. In addition, bioassessments do not identify causes of impairment. My thesis research had two general goals: 1) develop indices to measure biological condition in Nevada streams and 2) quantify relationships between land uses, stressors, and biological condition to infer likely causes of degradation.

I developed two types of biological indices based on stream invertebrates to measure biological condition and detect degradation in Nevada streams: (1) an index of taxonomic completeness, and (2) a multimetric index (MMI) that aggregates several attributes of invertebrate assemblages. The Nevada index of taxonomic completeness was relatively imprecise compared with indices developed for other regions, which I argue is a consequence of

the spatial isolation of aquatic habitats in arid regions. In contrast, the Nevada MMI was more precise than most previously developed indices, likely the result of using models to reduce variation in assemblage attributes along natural environmental gradients. Index values were positively related to how evenly individuals were distributed among taxa in a sample. Adjustments of index scores for sample evenness increased index precision and also altered inferences of biological impairment for specific sites.

I also developed models to relate measures of biological condition, instream stressors, and human land uses. Coupled with known causal mechanisms, associations between stressors and biological degradation can be used to infer likely causes of impairment. Total dissolved solids and metal contamination were the stressors most strongly associated with biological impairment. Stressors associated with degraded biological condition were also associated with land uses like agriculture, urbanization, and mining. The tools I developed to detect biological degradation and identify important stressors and their potential sources should help managers target conservation and restoration efforts and improve their ability to protect freshwater resources.

ACKNOWLEDGMENTS

I am very privileged to have had the opportunity to work on this project with the help and support of numerous mentors, peers, and friends. I would like to thank my advisor, Chuck Hawkins, and my committee members, Patrick Belmont and Richard Cutler, who have helped me to complete and improve my project as well as to help me become a better researcher. My labmates, Ryan Hill, John Olson, Robin Jones, Ellen Wakeley, and Nora Burbank, were all very helpful, and I am grateful for their support and friendship. Thank you to my friends and family, especially Natalie, for your continual support and encouragement. Finally, thank you to all of the friends I have made over the past two years for making my experience here so enjoyable.

Jacob J. Vander Laan

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The goal of the Clean Water Act (1972) is to "restore and maintain the chemical, physical and biological integrity of the Nation's waters." This goal has been a driving force behind research focused on the assessment, diagnosis, and restoration of stream ecosystems. In particular, there has been a strong focus on defining and measuring biological integrity and identifying biological impairment. Much of the effort to identify impairment has focused on bioassessment, which is often based on comparisons between the biota observed at a site and those that would be expected to occur in the absence of human disturbance (Cao and Hawkins 2011). Bioassessments have proven to be effective means to detect biological degradation in stream ecosystems, but do not diagnose the stressors responsible for degradation. To protect and restore the biological integrity of streams we must be able to both identify impairment and determine likely causes of degradation.

The ability to detect biological impairment is limited by the performance of bioassessment indices, and although index performance can vary widely, the sources of this variation are often unclear. It is also generally unclear how bioassessment indices will perform under different environmental settings. Most indices have been developed for mesic regions, and we know little about index performance in arid environments with isolated aquatic habitats. Most evaluations of index performance and attempts to improve it have focused on the types of biological information included in an index and the way that natural biotic variation is accounted for within the index (e.g. Cao et al. 2007, Van Sickle et al. 2007, Van Sickle et al. 2010, Hawkins et al. 2010). However, few studies have examined ways that natural physical and biological

factors may influence index performance, or sought to test index performance in arid regions with isolated aquatic habitats.

Although bioassessments can detect impairment, they do not identify potential causes of degradation. Observed biological degradation is generally assumed to be caused by one or more anthropogenic stressors, but these assumptions are often untested (Sloane and Norris 2003). Furthermore, the responsiveness of indices to specific stressors is generally unknown (Freund and Petty 2007). Models that link bioassessment index scores to known stressors and land uses would improve our understanding of index responsiveness to anthropogenic disturbance and help to identify important stressors and their sources.

Many types of stressors are known to cause biological degradation in streams, but diagnosing specific stressors responsible for degradation at individual sites is problematic. Quantifying relationships between stressors and biota can be complicated by the confounding of natural and anthropogenic gradients and the potential for complex interactions among stressors (Allan 2004, Townsend et al. 2008). Because cause and effect between stressors and biota can rarely be inferred from single studies, inferences regarding the most likely causes of degradation must often rely on a weight of evidence approach (e.g. Suter et al. 2010, Allan et al. 2011, Norris et al. 2012). In these approaches, causation of degradation can be most confidently attributed to a stressor when a causal mechanism has been established and there is a consistent and strong association between the hypothesized cause and effect (Suter et al. 2010, Allan et al. 2011, Norris et al. 2012).

To restore and maintain the biological integrity of stream ecosystems, managers must be able to both detect biological degradation and identify its likely causes. My thesis work focused on developing tools that can be used to measure biological condition and infer the likely

causes of impairment in Nevada streams, a highly arid, but also environmentally heterogeneous state. The unique environment of Nevada provided an interesting challenge for developing stream bioassessment indices and provided an opportunity to (1) assess the applicability of bioassessment indices to arid regions with high environmental heterogeneity and isolated aquatic habitats and (2) examine the effects of natural environmental and biological factors on index performance. In addition, the paired collection of invertebrate and stressor information allowed us to examine links between anthropogenic land uses, instream stressors, and biological condition and to identify important stressors associated with degradation. The tools I developed to measure biological condition, detect degradation, and identify important stressors and their potential sources should help managers better target conservation and restoration efforts and improve their ability to protect or restore freshwater resources.

Literature Cited

ALLAN, J.D., L.L. YUAN, P. BLACK, T. STOCKTON, P.E. DAVIES, R.H. MAGIEROWSKI, AND S.M. READ. 2011. Investigating the relationships between environmental stressors and the stream condition using Bayesian belief networks. Freshwater Biology doi: 10.1111/j.1365-2427.2011.02683.x.

ALLAN, J. D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology Evolution and Systematics 35:257-284.

CAO, Y., AND C. P. HAWKINS. 2011. The comparability of bioassessments: a review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680-701.

CAO, Y., C. P. HAWKINS, J. OLSON, AND M. A. KOSTERMAN. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26:566-585.

FREUND, J. G., AND J. T. PETTY. 2007. Response of fish and macroinvertebrate bioassessment indices to water chemistry in a mined Appalachian watershed. Environmental Management 39:707-720.

HAWKINS, C. P., Y. CAO, AND B. ROPER. 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066-1085.

NORRIS, R.H., J.A. WEBB, S.J. NICHOLS, M.J. STEWARDSON, AND E.T. HARRISON. 2012. Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshwater Science 31:5-21.

SLOANE, P. I. W., AND R. H. NORRIS. 2003. Relationship of AUSRIVAS-based macroinvertebrate predictive model outputs to a metal pollution gradient. Journal of the North American Benthological Society 22:457-471.

SUTER, G. W., II, S. B. NORTON, AND S. M. CORMIER. 2010. The science and philosophy of a method for assessing environmental causes. Human and Ecological Risk Assessment 16:19-34.

TOWNSEND, C. R., S. S. UHLMANN, AND C. D. MATTHAEI. 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45:1810-1819.

VAN SICKLE, J. 2010. Correlated Metrics Yield Multimetric Indices with Inferior Performance. Transactions of the American Fisheries Society 139:1802-1817.

VAN SICKLE, J., D. P. LARSEN, AND C. P. HAWKINS. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. Journal of the North American Benthological Society 26:319-331.

CHAPTER 2

ASSESSING BIOLOGICAL CONDITION OF ARID ZONE STREAMS: EFFECTS OF MODELING, SPATIAL ISOLATION, AND SAMPLE EVENNESS[1]

Abstract

In this study, we assessed the performance of two biological indices (a multimetric index and an O/E index of taxonomic completeness) that we developed for stream invertebrate assemblages in the arid state of Nevada, USA. Our primary objectives were to: (1) develop a general method to determine if assessed sites are represented by the population of reference sites; (2) develop indices that are insensitive to natural environmental gradients; (3) develop a robust method to select metrics for inclusion in MMIs that ensures maximum independence of metrics; and (4) assess the effects of site isolation, beta diversity, and sample evenness on index accuracy and precision. We developed a nearest-neighbor approach to determine which assessed sites occurred within the environmental space of the reference site network. This approach appears robust and equally applicable to all biological indices. Random Forest modeling effectively accounted for natural biological variability in both the multimetric and O/E indices and resulted in improved index performance over null models. Principal components analysis identified six statistically independent axes of variation among 31 candidate assemblage-level metrics. We then used one metric from each axis to produce a precise, responsive, and ecologically robust MMI. In contrast, the O/E indices we developed were relatively imprecise compared with O/E indices developed for other regions. This imprecision was the consequence of low predictability in local taxa composition, which was likely associated with the relatively high spatial isolation of aquatic habitats within arid regions. Estimates of taxa

[1] Coauthor: Charles P. Hawkins

richness and both index scores were positively associated with the evenness in the distribution of individuals among taxa within samples. This relationship can potentially compromise inferences regarding biological condition, and post-hoc adjustments for the effects of evenness on index scores might be desirable. Further improvements in the performance and interpretation of biological indices will require simultaneous consideration of the effects of incomplete sampling on characterization of biological assemblages and the natural processes that influence community assembly.

Introduction

Stream bioassessments are generally based on one of two types of biological indices: multimetric indices (MMIs) and observed to expected (O/E) taxa ratios (Cao and Hawkins 2011). An MMI aggregates several measures of invertebrate assemblage attributes (Karr and Chu 1997). The types of attributes used in MMI's often include measures of taxonomic richness, biological diversity and assemblage composition (Stoddard et al. 2008). Individual metrics that differentiate between reference and degraded condition are standardized and aggregated into a single measure of biological condition. O/E indices assess the taxonomic completeness of a site by comparing observed and expected taxa lists (e.g. Wright 1995, Hawkins 2006). The taxa expected at a site are predicted by identifying relationships between taxonomic composition and environmental gradients at pre-defined reference sites, and biological condition at assessed sites is measured as the number of expected taxa (E) that are observed (O) at a site, usually expressed as a ratio, O/E (Wright 1995). Benthic macroinvertebrate (BMI) based bioassessment indices have become increasingly important and widely used tools for assessing freshwater resources, but index performance varies widely, and the sources of this variation are not well known.

Several challenges must be overcome when developing robust MMIs and O/E indices, and the ways researchers address these challenges can influence index performance and interpretation. Accurate bioassessments require a comparison of the biota observed at a site with an estimate of that site's biological potential. Sites with similar natural environmental settings are considered to have similar potential. Accurate determination of biological expectations requires that the biological potential of any assessed site is adequately represented within the pool of reference sites. For O/E indices, Moss et al. (1987) developed a statistical test to determine if the combination of specific values of environmental attributes used to predict reference condition biota (E) at an assessed site occur within the pool of reference sites. However, their method, based on the discriminant functions used to predict E in the RIVPACS model, is not generally applicable to other types of indices. It would be useful to have a general approach to identifying sites that lack appropriate reference sites that could be used with all indices. MMIs have generally focused on ensuring that some minimum set of reference sites occurs within all regions or typologies, but how well those reference sites represent environmental gradients within regions has typically not been considered. A more broadly applicable method for identifying sites whose environmental conditions are not represented within the set of reference sites would help ensure the accuracy of a variety of bioassessment indices when applied to sites of unknown condition.

One of the biggest challenges for any bioassessment is accounting for naturally occurring spatial and temporal variation in assemblage composition (Hawkins et al. 2010b). Failure to adequately account for natural variation in assemblage composition can result in confounding of natural and human-caused effects on assemblages and lead to inaccurate assessments of biological condition. Another challenge involves improving the sensitivity of

indices so that both moderate and severe alteration can be detected. For example, some O/E developers have sought to improve index performance by excluding taxa with low probabilities of capture. Some studies have shown that excluding locally rare taxa improves O/E index performance (e.g. Hawkins et al. 2000b, Ostermiller and Hawkins 2004, Van Sickle et al. 2007), but exclusion of rare taxa does not always improve O/E index performance (Van Sickle et al. 2007). Redundant metrics can degrade MMI performance (Van Sickle 2010), and MMI developers have attempted to improve index performance by selecting statistically and biologically independent metrics for inclusion in MMIs. However, the routinely used methods for selecting metrics may not ensure metric independence or improve MMI performance and interpretability (Van Sickle 2010). Incomplete sampling can also affect index performance (Ostermiller and Hawkins 2004, Cao and Hawkins 2005). Bioassessments are based on samples that are assumed to accurately represent stream assemblages. However, both sample count and sample evenness may affect their representativeness, repeatability, and comparability. The development of robust indices requires that we thoroughly address these issues.

Metric redundancy is generally considered to compromise MMI performance (Van Sickle 2010), and index developers have tried to minimize redundancy in several ways. Approaches for identifying redundant metrics include setting a maximum value for the allowed correlation between metrics in an MMI (Barbour et al. 1999, Hering et al. 2006, Stoddard et al. 2008), using best professional judgment to classify metrics into ecologically independent categories (Hering et al. 2006, Stoddard et al. 2008), and objectively clustering metrics based on their correlations (Cao et al. 2007). Typically in these approaches, a single or equal number of metrics from each set of redundant metrics is selected for the final index. However, these approaches do not necessarily ensure statistical or biological independence of metrics and require the application

of an arbitrary correlation coefficient cutoff. Van Sickle (2010) demonstrated that the maximum correlation between two metrics in an MMI has little or no effect on index performance and instead suggests that minimizing the mean correlation between metrics is a more effective method of increasing index performance. An objective method for reducing redundancy among metrics could help ensure the development of ecologically and statistically robust MMIs.

Failure to account for the effects of natural variability on the abundance and distribution of biota can lead to confounding of natural and anthropogenic effects on assemblages. This confounding can lead to biased assessments and result in increased type I and type II errors of inference, i.e., a site being assessed as impaired when it is not and vice versa. O/E indices and MMIs generally differ in how they account for natural variability in assemblages and define biological expectations. Biological expectations for O/E indices are set by modeling natural variability among assemblages and predicting site-specific expectations of the probability of observing individual taxa. MMIs have traditionally relied on some type of regionalization to account for the natural variation of metric values among reference sites and thus set expectations for biological condition at assessed sites (Barbour et al. 1999). However, these regionalizations often account for little of the biological variation that occurs between sites (Hawkins et al. 2000a, Hawkins et al. 2010). Recently, some MMI developers have also used modeling techniques to adjust metric expectations for natural gradients (e.g. Baker et al. 2005, Pont et al. 2006, Cao et al. 2007, Hawkins et al. 2010a), which can increase MMI accuracy and precision (e.g. Cao et al. 2007, Hawkins et al. 2010a).

Both O/E indices and MMIs are based in whole, or in part, on measures of taxonomic richness. Therefore, any sample property, such as sample evenness, that affects estimates of richness may affect index performance. For example, the taxonomic richness of stream

invertebrate samples typically increases asymptotically with the number of individuals in a sample (Vinson and Hawkins 1996). However, this asymptote will be approached more rapidly in highly even assemblages than in less even ones (Gotelli and Colwell 2001). Therefore, differences in sample evenness may lead to differences in estimates of sample richness and composition. This dependency of estimated richness on evenness will be particularly problematic for the small, fixed-count samples commonly used in bioassessments (Cao and Hawkins 2005). Differences in sample evenness could therefore increase bias and decrease precision of O/E indices. Because MMIs are often comprised of richness attributes, MMI scores may also be affected by sample evenness.

Bioassessment index performance varies widely, but the causes of this variation are often unclear. Decisions about how to account for natural biological variability, how to minimize the effects of sampling variability, and what biological information should be included in an index can all affect index performance. However, the environmental and biological properties of the study area may also influence index performance. Environmental heterogeneity, habitat isolation, and the magnitude of taxa turnover among sites (i.e., beta diversity) may affect taxa predictability and index performance.

In this study, we developed an O/E index and a modeled MMI for Nevada (NV) streams. NV's generally arid, but heterogeneous, landscape provided an excellent opportunity to explore regional sources of variation in index performance by comparing indices developed for NV with indices developed for other regions. Our primary objectives were to: (1) develop a general method to determine if assessed sites are represented by the population of reference sites; (2) develop both MMI and O/E indices that are insensitive to natural environmental gradients; (3) develop a robust method to select metrics for inclusion in MMIs that ensures maximum

independence of metrics; and (4) assess the effects of site isolation, beta diversity, and sample evenness on index accuracy and precision.

Methods

Study area and data

We used a NV Department of Environmental Protection (NDEP) dataset consisting of macroinvertebrate samples collected at 500 stream sites throughout NV and nearby surrounding areas to develop MMI and O/E indices for the state (Fig.2-1). Data included samples from 415 sites of unknown biological condition collected by the NDEP and samples from 85 sites previously determined to be in reference condition that were sampled by either the Western Center for Monitoring and Assessment of Freshwater Ecosystems (WMC, Utah State University) or the U.S. Environmental Protection Agency.

The state of NV lies mostly within the Great Basin. This region is environmentally heterogeneous with environments that range from desert to montane. The Basin and Range terrain produces extreme climactic variability, and cool, wet climates are restricted to isolated high-elevation habitats. These high-elevation habitats are sometimes referred to as sky islands, and despite their small size and isolation, support most of the aquatic habitats and much of the total biodiversity of the region (Chambers et al. 2008).

Watershed delineation and predictor variable extraction

We used 27 different environmental variables as possible predictors of macroinvertebrate assemblages in our models (Appendix A). We delineated watersheds with the Multi-Watershed Delineation Tool (Chinnayakanahalli et al. 2006). This tool uses digital elevation grids to identify drainage divides and delineate watersheds. From these delineations

we then extracted candidate predictors. These predictors included characterizations of the climate, soils, geology, hydrology, topography, and water chemistry of each site and watershed. We extracted mean, minimum and maximum values from thirty-year average (1971-2000) temperature and precipitation information derived from grids produced by the PRISM climate group (Daly et al. 2008). We used available geologic (Reed and Bush 2001) and soil data (Wolock 1997) to characterize soils and geology within watersheds and at sample locations. We characterized two aspects of the hydrologic regime: (1) mean base flow index (Wolock 2003) and (2) hydrologic stability calculated as the minimum mean monthly discharge divided by the maximum mean monthly discharge interpolated from USGS gauging stations. We characterized watershed topography with three elevation measures (mean, minimum and maximum elevation within a watershed), as watershed size, and by a measure of watershed slope calculated as the change in elevation in the watershed divided by the maximum flow length. We also characterized near-site topographic variability as the coefficient of variation of the elevations occurring within a radius of five 30-meter digital elevation model cells from the sample site (about 80 cells). Finally, we used predicted natural stream water conductivity to characterize differences in naturally occurring water chemistry among sites (Olson and Hawkins 2012).

Reference and most degraded site selection

We used an approach similar to that described by Herlihy et al. (2008) to screen potential reference sites based on watershed land use characteristics and potential human impacts. We used land cover information obtained from the National Land Cover Database (NLCD) produced by the Multi-Resolution Land Characteristics Consortium (Homer et al. 2007) to calculate the percent of area in each watershed that was classified as agricultural or urban land use. We also calculated percent of the watershed within 3 km upstream from the sample

site that was classified as agricultural or urban land use. Additionally, we used a National Pollutant Discharge Elimination System (NPDES) GIS coverage to identify major point-source discharges into each stream.

We stratified watersheds into three size categories and selected potential reference sites from each category. We used this stratification to ensure that we included reference sites that spanned the range of stream sizes and environmental conditions present in NV. Differences in the degree of watershed alteration associated with watershed size required that we adopt less stringent thresholds for the larger streams. For each size category, we set selection thresholds for each of the screening variables, with the most rigorous standards set for small watersheds and progressively less stringent standards for medium and large watersheds (Table 2-1). Thresholds were established by examining the distributions of indicators of human disturbance within each watershed size category and selecting the lowest possible threshold that retained enough sites for robust modeling. We used these thresholds to select the least altered watersheds available in each size category as potential reference sites.

For small and medium watersheds, we also used Google Earth to look for other possible human modifications not detected in the previous screening such as the presence of dams, mining, heavy recreational use, or extensive roads in the watershed. We excluded watersheds for which satellite imagery showed significant signs of degradation. We performed similar screening for large watersheds, but because all large watersheds have some form of degradation within them, we focused on reach-scale conditions at each site.

We used available water chemistry information to identify possible degradation not detected through the GIS and satellite imagery screening. We excluded sites with conductivity, nutrient concentrations, or metal concentrations that were distinct outliers relative to the

chemical characteristics of other reference sites. Finally, NDEP personnel reviewed the list of potential reference sites, and sites were added or removed based on their recommendations.

In addition to identifying reference sites, we selected 60 most degraded sites for index development and assessment. We used GIS, Google Earth and available water chemistry to identify sites with greatest apparent human-caused environmental degradation. No formal thresholds were established, and no biological information was used to identify the most degraded sites.

Identifying sites outside of reference site environmental space

We developed a nearest-neighbor based approach to identify sites with environmental characteristics that were outside of the environmental space defined by the reference sites. We used environmental factors generally known to be associated with spatial variation in stream community structure and composition and confirmed by both our MMI and O/E models (see below) to define ecologically relevant environmental space. We transformed variables as necessary to approximate normal distributions, performed PCA with varimax rotation on the variables to identify independent axes of variation, and selected the highest loading variable from each axis to represent all correlated variables on an axis. We then standardized these representative variables by scaling between the minimum and maximum values observed in the reference data. We used the standardized variables to calculate Euclidean multivariate distances between each reference site and all other reference sites. We then calculated the average distance of each reference site to the 10 nearest other reference sites and used the 90^{th} percentile of this distribution as a threshold for defining if a new site was outside of reference site environmental space. To apply this test to new sites, we calculated the average distance of each new site to the 10 nearest reference sites and flagged a new site as an outlier if the

average distance exceeded the 90th percentile threshold defined by the distribution of 10-nearest neighbor reference site distances.

BMI data

We applied a common taxonomy and random subsampling procedure to all BMI samples. All taxa in the NDEP dataset were aggregated into unique operational taxonomic units (OTU). The use of OTUs allows a common taxonomy to be applied to all samples and ensures consistent taxonomic resolution between samples (Hawkins et al. 2000b). To correct for possible differences in sampling effort between samples, we randomly re-sampled all macroinvertebrate samples to fixed counts of 300 individuals (Ostermiller and Hawkins 2004). For those samples with fewer than 300 individuals, all individuals were retained. The fixed-count, OTU standardized, samples were used for both MMI and O/E development.

MMI development

For MMI development, we started with over 100 possible macroinvertebrate metrics. Based on a comprehensive literature review, we reduced this list to 31 commonly used or potentially useful metrics (Table 2-2). Because the values of at least some metrics used in an MMI are likely to vary along environmental gradients, we used empirical models to estimate the metric values expected at sites as a function of natural environmental attributes. Using reference site data, we built Random Forest (RF) models (Cutler et al. 2007) for each metric. If natural environmental gradients were associated with > 10% of the variation in metric values at reference sites, we extracted residual values by subtracting the predicted value from the observed value. These residuals are a measure of metric variability after adjusting for natural environmental gradients. The residual values were then used in MMI development.

We used principal components analysis (PCA) and t-statistics to select metrics for inclusion in the MMI. We performed PCA with varimax rotation on the adjusted metrics to identify statistically independent axes of variation among the 31 candidate metrics. We then conducted t-tests on the mean metric values from samples at reference and most degraded sites to assess the ability of individual metrics to differentiate between reference and degraded condition. We selected the metric from each axis of variation that showed the greatest differentiation (largest absolute t-value) between reference and degraded sites to include in the MMI. Selecting one metric from each axis maximized the statistical independence of metrics used in the MMI. We then rescaled each metric following Cao et al. (2007) to a score between 0 and 100. Metrics that decreased with disturbance were rescaled as:

$$100 \times \left(\frac{site\ value - min}{max - min} \right)$$

and metrics that increased with disturbance were rescaled as:

$$100 \times \left(\frac{site\ value - max}{min - max} \right)$$

where the minimum (min) value is the 5th percentile of metric values at highly degraded sites, and the maximum (max) value is the 95th percentile of metric values at reference sites. The use of the 5th and 95th percentiles reduces the impacts of possible outliers, and has been shown to produce MMIs with high responsiveness, but low variability (Blocksom 2003, Stoddard et al. 2008). We calculated the final MMI by summing the scores of all seven metrics from a sample, then dividing by the number of metrics.

For comparison purposes, we also constructed a null MMI from the same 31 candidate metrics but without model adjustment. We used PCA with varimax rotation and t-statistics to select metrics for inclusion in the null MMI and followed the same rescaling and MMI calculation procedure as before. To allow comparison between the MMI and O/E indices, we also calculated standardized MMI scores by dividing raw MMI scores by the mean of reference site scores so that reference site scores were centered on one.

O/E index development

We developed O/E indices based on relationships between environmental gradients and taxa observed at reference sites. We clustered reference sites based on pair-wise Sorensen distances in assemblage composition. For clustering purposes only, we removed taxa observed at less than 10% of sites. We then developed an RF model to predict cluster membership from the natural environmental characteristics of each site. We used the RF model to predict the probabilities of cluster membership for any given site. We used the probabilities of cluster membership and taxon occurrence frequencies within reference site clusters to predict taxon-specific probabilities of capture (P_c) (Moss et al. 1987). We calculated expected richness at a site as the sum of all individual taxon P_c values at a site greater than a set threshold. Observed richness was calculated as the number of those taxa with P_c greater than the set threshold that were observed in the fixed-count samples. We assessed the effects of excluding locally rare taxa by calculating O/E scores based on P_c values ranging between 0 and 0.6 at 0.05 intervals and assessing the precision and sensitivity of these indices. We used RF out-of-bag (OOB) predictions to estimate expected richness of calibration sites. We calculated O/E as the proportion of expected taxa (E) that were observed (O) in a sample. For comparison with these indices, we also calculated null O/E indices in which probabilities of capture of individual taxa are assumed

to be identical across all sites (Van Sickle et al. 2005). The precision of a null index represents theoretical lower boundaries for O/E index precision (Van Sickle et al. 2005).

Inferring biological condition

We used one-sided non-central interval and equivalence tests to define thresholds between reference and impaired biological condition for the final MMI and O/E indices. We established thresholds for both equivalence to and dissimilarity from the reference distribution at 95% confidence. Scores that fell between these two thresholds were considered undetermined. This process resulted in three condition classes: reference, impaired, and undetermined.

Evaluating and comparing indices

We applied the MMI and O/E indices to 60 samples from heavily degraded sites and a set of 416 samples from sites of unknown condition (test sites) and evaluated the performance of each index. We followed procedures similar to those in Hawkins et al. (2010a) and evaluated the indices based on four measures of performance: bias, precision, responsiveness, and sensitivity. We used additional RF models to assess index bias by determining if any variation in reference site index scores were still associated with environmental gradients. We measured bias as the percent of variation in reference site index scores explained by a set of 11 natural environmental predictors including measures of temperature, precipitation, elevation, hydrologic stability, and watershed slope and size. We measured index precision as the coefficient of variation of reference site scores for each index. We used two measures of index responsiveness: the difference in mean index scores between reference and most-degraded sites, and Student's t values for the differences between reference and most-degraded index

scores. We evaluated index sensitivity as the percent of test samples inferred as being in non-reference condition; i.e., sites assessed by an index as either degraded or undetermined. For this evaluation, we considered samples with scores below the reference score equivalence threshold to be in non-reference condition. We also compared the MMI and O/E indices by calculating correlation coefficients and slopes of regression lines between MMI and O/E index scores for test samples and testing whether the slopes were significantly different from 1. We assessed six different indices: O/E indices with $P_c > 0$ and $P_c \geq 0.5$ (O/E-0 and O/E-5), null O/E indices for the two modeled indices (O/E-0-null and O/E-5-null), the modeled MMI, and the null MMI with no RF adjustments. We also compared the O/E indices we developed for NV to a relatively precise index (CV = 0.13 $P_c \geq 0.5$, CV = 0.19 $P_c > 0$) developed for Colorado (CO) mountain streams (Hawkins 2009) and examined possible reasons for differences in index precision, particularly differences in beta diversity among NV and CO. Because of the wide array of beta diversity indices and uncertainty as to which indices appropriately characterize beta diversity (Koleff et al. 2003), we did not calculate specific beta diversity index values. Instead, we assessed beta diversity for both NV and CO by calculating the number of shared and unique taxa between all pairs of reference samples within each state. We expressed these shared and unique taxa as box plots of all pair-wise comparisons and as ternary plots (Koleff et al. 2003). The ternary plots were constructed by plotting each pair-wise comparison where coordinates are defined by the proportion of shared taxa between samples and the proportions of unique taxa of each sample.

Effects of sample evenness on indices

To determine if sample evenness affected index performance, we regressed both sample richness and index scores on sample evenness. We measured sample evenness as Hill's evenness calculated as Shannon diversity divided by the natural log of richness. We developed

linear regression models relating index scores to sample evenness for pooled reference and test sites and used the regression to apply a Post hoc adjustment to index scores. We calculated adjusted index scores as: $(S - S_{pred} + 1)$, where S is the original index score (O/E or standardized MMI score) and S_{pred} is the index score predicted from the evenness regression model. We validated by randomly withholding one third of reference samples from the regression models and assessed the effects of the adjustment as change in CV of adjusted validation sample scores. For comparison, we performed the same analysis and adjustment on the CO mountain stream O/E index.

Results

Reference site selection

Of the 165 reference quality sites that we used for index calibration (Fig.2-1), 133 sites had small watersheds, 25 sites had medium sized watersheds, and 7 sites had large watersheds. Eighty reference sites were identified from the NDEP dataset and 85 were added to the reference site pool from the WMC and USEPA datasets.

Identifying sites outside of reference site environmental space

The PCA on all predictors used in either the MMI and O/E indices identified three axes of environmental variation, and we selected watershed area (log transformed), mean watershed elevation, and watershed maximum annual precipitation (log transformed) to represent these axes. When applied to the 416 test samples, 19 sites (49 samples) were flagged as environmentally different from the reference sites. Bivariate plots showed that most of these flagged sites were outside or on the fringes of the cloud of reference site environmental space and were generally larger, drier, lower-elevation sites (Fig.2-2).

MMI development

Of the 31 metrics considered for the MMIs, natural gradients explained greater than 10% of the variation in all but two metrics: Shannon diversity and Simpson diversity (Table 2-3). These metrics were therefore not adjusted for MMI development. The PCA on 31 adjusted metrics resulted in seven statistically independent axes of variation (Table 2-4). These axes represented different aspects of assemblage structure: richness, composition, diversity/dominance, and tolerance. Although %EPT had the largest absolute t-value in axis two, we chose to replace %EPT with %EPHEM in the MMI to avoid taxonomic redundancy with %PLEC. The modeled MMI consisted of seven metrics: INS-RICH, %EPHEM, SHDIVER, %CF, %PLEC, NON_INS-RICH, and CLING-RICH. Correlations between modeled metrics ranged between 0.016 and 0.71 with a mean correlation of 0.17. The maximum correlation of modeled metric scores at reference sites (0.71) was between INS-RICH and SHDIVER. In contrast to the modeled MMI, six axes of variation were identified for the unadjusted metrics. The metrics included in the null MMI were: INS-RICH, HBIINT1-RICH, SHDIVER, %HYDRA, EPHEM-RICH, and LLT-RICH. Correlations between non-modeled metrics ranged between 0.003 and 0.63 with a mean correlation of 0.20. The maximum correlation of non-modeled metric scores (0.63) was between INS-RICH and SIDIVER.

O/E index development

We identified eight reference site clusters for O/E index development with group sizes ranging between 8 and 29 sites. Seven environmental predictors were used in the RF model to predict class membership: watershed area, watershed slope, mean watershed elevation, watershed-level hydrostability, watershed long-term maximum annual precipitation, watershed long-term maximum annual temperature, and predicted conductivity. This model classified

62.4% of reference sites correctly, and within-class prediction errors ranged between 21% and 67%.

The P_c threshold used in O/E significantly influenced index precision, sensitivity, and expected richness at reference sites. As the P_c threshold used in the O/E index increased, index precision decreased (Fig.2-3). The standard deviation increased with P_c for both modeled and null indices. Index sensitivity peaked at $P_c \geq 0.2$, then generally declined as P_c was increased. Finally, increasing P_c significantly reduced the expected richness at reference sites with average expected richness at reference sites decreasing from 32.2 to 8.7 with an increase in P_c from 0 to 0.5 (Fig.2-3).

Evaluating and comparing indices

In all cases, a modeled index was more precise than its null counterpart. Modeling natural gradients eliminated index bias, with natural gradients accounting for < 1% of the variation of reference site index scores (Table 2-5). In contrast, the three null indices all showed significant bias. Reference site scores for the three null indices were most heavily influenced by watershed size. These indices tended to score samples from larger watersheds lower than small watersheds. Clear, but less pronounced, biases were also observed for other gradients, with each null model demonstrating its own specific biases. The modeled MMI was the most precise, sensitive, and responsive index (Table 2-5).

O/E and MMI scores for test sites were strongly correlated with one another (Fig.2-4), and all indices differentiated between reference and most degraded sites (Table 2-5). Test site scores from the modeled MMI were more correlated with test site scores from O/E-0 (r=0.79) than with O/E-5 (r=0.56). There was a stronger correlation between O/E-0 and the modeled MMI test site scores than between O/E-0 and O/E-5 (r=0.52). The regression slopes for the

modeled MMI versus values for both O/E indices were significantly different from one ($p<<0.001$ for both), but the slope of the regression line for O/E-0 versus O/E-5 was not significantly different from one ($p = 0.33$).

Clear differences existed in both index precision and beta diversity between the CO and NV data. The CO O/E indices were more precise for both $P_c > 0$ and $P_c \geq 0.5$ (Table 2-5). At $P_c > 0$, both observed and expected richness were higher in NV than in CO, but this pattern was reversed at $P_c \geq 0.5$ (Fig.2-5). In addition, significant differences existed in beta diversity between the NV and CO samples. In pair-wise comparisons, samples from NV reference sites tended to have fewer shared taxa than unique taxa, whereas the CO samples tended to have similar numbers of shared and unique taxa (Fig.2-6). Similarly, the ternary plots showed that the CO samples generally had higher proportions of shared taxa than the NV assemblages (Fig.2-6).

Effects of sample evenness on index performance

Sample evenness was associated with sample richness and index scores in both the NV and CO data. In the NV data, sample evenness ranged between 0.32 and 0.90 with a mean of 0.72. We found no association between sample evenness and natural environmental gradients. Evenness at reference sites (mean = 0.75) was slightly higher than at test sites (mean = 0.72) and degraded sites (mean = 0.65) (Fig.2-7). Differences in sample evenness between reference site and test or degraded site evenness were both statistically significant (t-tests, $p = 0.001$ and $p < 0.0001$, respectively). In CO, sample eveness ranged between 0.38 and 0.85 with a mean of 0.74. Similar to NV, we found no association between sample evenness and environmental gradients in CO, and evenness at reference sites was slightly higher (mean = 0.74) than at a set of stressed sites (mean = 0.58) (Fig.2-7). This difference was also statistically significant ($p<0.001$). The relationship between evenness and richness was stronger in NV ($r^2=0.33$) than in CO ($r^2=0.08$).

With the exception of the CO O/E-5 index, index scores for both NV and CO were also positively related to sample evenness ($p<0.05$). Evenness was most strongly associated with the NV MMI scores. Relationships between evenness and index scores were stronger in NV than in CO (Table 2-6). Post hoc adjustments of index values for differences in evenness resulted in small improvements in precision in all cases, with the greatest improvement observed in the NV MMI. Adjustments for evenness resulted in slight decreases in index sensitivity for the NV O/E indices, and a small increase in sensitivity for the NV MMI. Adjustments also altered inferences of biological condition from each index for specific NV test sites (Table 2-7).

Discussion

Methods of bioassessment index development and the environmental and biological properties of a study area can both affect the index performance. To develop robust indices, we need to account for natural variability in assemblages, include appropriate biological information, and reduce the effects of incomplete sampling on estimates of richness and composition. We also need to understand how index performance might be affected by differences among biogeographic regions in the processes responsible for community assembly and maintenance. This study provided new insights regarding each of these issues.

Accounting for natural variability in assemblages

Modeling resulted in increased precision, sensitivity, and responsiveness over null indices. It also eliminated bias associated with natural gradients. Although modeling did not increase the sensitivity of the O/E-5 index, it did eliminate bias, which would have been manifested as apparent sensitivity to degradation in the null index. In particular, our results suggest that modeling metrics for MMIs can greatly improve the precision and performance of

MMIs which should lead to more accurate assessments of biological condition. Our results are consistent with the findings of previous studies (e.g. Cao et al. 2007, Hawkins et al. 2010a) in suggesting that modeling to account for natural variability in biota should improve bioassessment index performance and decrease type I and type II errors of inference.

Although modeling can improve bioassessment index performance, when the pool of reference sites does not include the environmental characteristics of an assessed site, the models may not accurately account for natural variability in assemblages. This failure to account for natural variability can lead to inaccurate expectations of biological condition and thus inaccurate assessments. The method that we developed to detect sites with environmental characteristics that are not within the environmental space of the reference site pool should help ensure accurate assessments regardless of the type of index used. Our use of the 90^{th} percentile of reference site distance distributions as a threshold for defining a new site as outside reference site conditions was arbitrary, but can be easily changed to accommodate specific indices and assessment goals. The strength of this method is that it is broadly applicable and can be readily adapted to any bioassessment.

Selecting biological information for use in indices

MMIs. – Metric independence should result in ecologically and statistically robust MMIs. Our approach to maximizing metric independence is similar to the clustering approach of Cao et al. (2007) in that, like cluster analysis, PCA groups metrics into statistically related axes. However, use of PCA ensured that the selection of one metric from each axis maximized the statistical independence of the metrics used in the final index. As suggested by Van Sickle (2010), our approach for selecting metrics focuses on reducing the average correlation between metrics in an MMI rather than setting an arbitrary maximum correlation coefficient. In addition,

our approach effectively separated metrics into different aspects of assemblage structure and organization, a general goal of MMIs (e.g. Barbour et al. 1999, Hering et al. 2006, Stoddard et al. 2008), and should ensure that measures of all important aspects of invertebrate assemblages are included in the MMI. However, this result is not necessarily general and should be further explored. Our approach also limits the difficulties and potential bias associated with *a priori* classifications of metrics based on professional judgment. However, the use of professional judgment may still be necessary at times. For example, we chose to replace % EPT with % Ephemeroptera to ensure compositional independence with % Plecoptera.

Selecting appropriate metrics for inclusion in an MMI is critical to producing an accurate and ecologically meaningful index. Hering et al. (2006) argued that a good metric should exhibit naturally low temporal and spatial variability. However, 29 of the 31 candidate metrics used for MMI development in this study exhibited high spatial variability associated with natural gradients, and selecting metrics with low natural variability may be unrealistic and eliminate useful metrics. In addition, Stoddard et al. (2008) argued that the signal to noise ratio of a good metric (a metric's variance among all sites compared to its variance between repeated visits to the same sites) must be high. A high signal to noise ratio implies metric values with high spatial variability among sites and conflicts with the idea that metrics with high spatial variability are inappropriate for MMIs. Finally, if natural variability of metrics is not accounted for before selecting metrics for an MMI, then the effects of anthropogenic stress will be confounded by natural gradients, and inferences regarding the true biological response to anthropogenic stress will be compromised. For instance, *Hydra* relative abundance, for which environmental gradients accounted for 22% of the variation, appeared to moderately discriminate between reference and stressed sites (t = -2.0) before modeling, but this response was reduced after

modeling (t = 1.0) (Table 2-4). In this case, not only was metric responsiveness to stress inflated, but the directionality of the response was also reversed after accounting for natural variation. Of the 29 metrics that were modeled with RF, 18 showed reduced responsiveness after accounting for natural gradients. Conversely, some metrics may reveal higher responsiveness to anthropogenic stress after modeling. In this dataset, the estimated responsiveness to stress of prey taxa richness, percent EPT, percent Ephemeroptera, and non-insect taxa richness all increased after accounting for natural variability. Modeling to account for natural variability ensures that the metrics included in an MMI are responding to anthropogenic degradation, and should improve overall MMI performance.

O/E indices. – Our results indicate that the effects of excluding locally rare taxa on O/E index performance are not yet fully understood. In contrast to Van Sickle et al.'s (2007) observation that excluding rare taxa tends to improve index precision, we found that excluding locally rare taxa in this dataset decreased index precision over much of the range of P_c. Because rare taxa are often less predictable than more common taxa, excluding them should increase index precision. However, increasing the P_c value used in an O/E index also decreases the magnitude of E, the number of taxa predicted to occur at any given site. As E decreases, the effect of a random non-detection of a taxon in a sample on the O/E score increases. This means that when E is small, the effects of random sampling error will likely be high, and index precision may be reduced. For instance, increasing Pc from 0 to 0.5 decreased average expected richness at NV reference sites from 32.2 to 8.7. This effect was smaller in the CO data with a decrease from 27.0 to 10.8. In some cases, the effect of improved predictability from excluding locally rare taxa may be countered by the increased effect of random sampling error. In general, the

optimal P_c value is likely to vary from region to region, and identifying the optimal P_c will require calculating O/E performance measures at several P_c values.

Regional differences in index performance

The NV O/E indices were relatively imprecise compared with the CO index as well as several indices developed for other regions. For instance, the SD of reference site O/E values based on $P_c \geq 0.5$ for the Mid-Atlantic Highlands (SD=0.18, Van Sickle et al. 2005), the Interior Columbia River Basin (SD=0.11, Hawkins et al. 2010a) and the states of California (SD=0.16, Hawkins et al. 2000b) and Oregon (SD=0.17, Van Sickle et al. 2006) are all more precise than the NV O/E-5 index (SD=0.26). In addition, both Yuan et al. (2008) and Hawkins (2009) observed that O/E indices for plains regions tended to be less precise than indices developed for mountainous regions. This decreased precision could result from either increased temporal variability in arid plains streams (e.g. higher frequency of drying events) or because environmental characteristics that are good predictors of assemblages in mountainous streams are not applicable in lowland streams. Because the NV O/E indices include both mountainous and lowland streams, either of these mechanisms could explain the decreased precision observed in NV. However, the precision of reference site O/E scores from only mountainous sites in NV (defined from the EPA level IV ecoregions, n=111) is nearly the same as that for the full indices (CV = 0.21 for $P_c > 0$ and CV = 0.22 for $P_c \geq 0.5$). We therefore suspect that the physical isolation of streams in NV probably results in reduced predictability of NV invertebrate assemblages. Where habitat isolation limits colonization, environmentally similar habitats will not necessarily contain similar assemblages – hence the high beta diversity values observed in NV. This interpretation is consistent with current metacommunity theory (e.g. Leibold et al. 2004). In environmentally heterogeneous regions, both beta and regional diversity have been shown to decrease with

increased connectivity and dispersal between habitats (Forbes and Chase 2002, Matthiessen et al. 2010). When dispersal is limited, the spatial distributions of dominant taxa are reduced and both beta and regional diversity are increased (Matthiessen et al. 2010). If dispersal is limited among NV streams, taxon presence will not be well predicted by environmental suitability alone. In contrast, predictions of taxon occurrence based on environmental suitability should be more accurate in regions with more connectivity, and indices developed for these regions are likely to be more precise. In general, the environmental and biological properties of a study area, such as habitat isolation and dispersal ability of taxa, likely affect the predictability of assemblages and may limit the precision of bioassessment indices.

Effects of sample evenness on index performance

Variations in sample evenness may reduce the performance of bioassessment indices derived from small, fixed-count samples. Because sample evenness can affect estimates of richness and composition, differences in sample evenness can bias and reduce precision of bioassessment indices. If bioassessments were based on census data, differences in assemblage evenness would be irrelevant to index performance. However, a census is impractical for invertebrate assemblages, and richness and composition must therefore be estimated from representative samples. Samples from assemblages with low evenness will underestimate richness relative to more even samples. This effect will be most problematic when sample sizes are relatively small, as is often the case in essentially all bioassessments (Cao and Hawkins 2005). As demonstrated by the differences between NV and CO (Table 2-6), the strength of the relationship between sample evenness and estimates of richness is likely to vary, but factors that affect evenness and the strength of the relationship between richness and evenness are unclear. Post hoc adjustments of index scores will be most beneficial where this relationship is

strong. The strong relationship between the NV MMI and sample evenness is likely the result of including four metrics that are likely to be affected by evenness (insect richness, Shannon diversity, non-insect richness, and clinger richness). If those individual metrics are all affected by sample evenness, then the effects of sample evenness may be amplified when they are combined in an MMI. Because MMIs often include several metrics that are likely to be affected by evenness, the effects of evenness may be especially problematic for MMIs. Our results indicated that low evenness may be associated with degradation in both NV and CO streams (Fig.2-6), but Cao and Hawkins (2005) showed that evenness can either increase or decrease with stress. Regardless of the sources of variation in evenness, we need to minimize its effect on estimates of index scores unless evenness is an explicitly defined assemblage attribute desired in an MMI. DNA-based methods may eventually allow us to enumerate taxa and estimate richness from very large samples (e.g. Pfrender et al. 2010, Hajibabaei et al. 2011) and hence reduce the dependency of richness estimates on sample evenness. However, until these methods are more practical, post hoc adjustments of O/E and MMI scores may be the only realistic way to minimize the effects of sample evenness on index values.

Implications for future bioassessments

To develop statistically and ecologically robust bioassessment indices, we need to thoroughly assess the performance of the indices we develop and strive to improve their accuracy and precision. In particular, researchers need to account for natural variability in biota, select appropriate biological information for use in indices, and minimize the effects of sampling variability and error. Previous studies provide some guidelines for addressing these issues, but more comprehensive approaches are needed that can allow us to scrutinize how both sampling artifacts and naturally occurring ecological properties affect index performance.

Literature Cited

BAKER, E. A., K. E. WEHRLY, P. W. SEELBACH, L. WANG, M. J. WILEY, AND T. SIMON. 2005. A multimetric assessment of stream condition in the northern lakes and forests ecoregion using spatially explicit statistical Modeling and regional normalization. Transactions of the American Fisheries Society 134:697-710.

BLOCKSOM, K. A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic highlands streams. Environmental Management 31:670-682.

BARBOUR, M.T., GERRITSEN, J., SNYDER, B.D., AND J.B. STRIBLING. 1999. Rapid bioassessment protocols for use in streams and wadable rivers: periphyton, benthic macroinvertebrates, and fish, second edition. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

CAO, Y., AND C. P. HAWKINS. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. Journal of Applied Ecology 42:954-965.

CAO, Y., AND C. P. HAWKINS. 2011. The comparability of bioassessments: a review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680-701.

CAO, Y., C. P. HAWKINS, J. OLSON, AND M. A. KOSTERMAN. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26:566-585.

CHAMBERS, J.C., DEVOE, N., AND A. EVENDEN. 2008. Collaborative management and research in the Great Basin — examining the issues and developing a framework for action. Gen. Tech. Rep. RMRS-GTR-204. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 66 p.

CHINNAYAKANAHALLI K, KROEBER C, AND HILL R. 2006. The multiwatershed delineation tool: GIS software in support of regional watershed analyses. Utah State University, Logan, UT.

CUTLER, D. R., T. C. EDWARDS, K. H. BEARD, A. CUTLER, AND K. T. HESS. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

DALY, C., M. HALBLEIB, J. I. SMITH, W. P. GIBSON, M. K. DOGGETT, G. H. TAYLOR, J. CURTIS, AND P. P. PASTERIS. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology 28:2031-2064.

FORBES, A. E., AND J. N. CHASE. 2002. The role of habitat connectivity and landscape geometry in experimental zooplankton metacommunities. Oikos 96:433-440.

GOTELLI, N. J., AND R. K. COLWELL. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379-391.

HAJIBABAEI, M., S. SHOKRALLA, X. ZHOU, G. A. C. SINGER, AND D. J. BAIRD. 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. Plos One 6(4): e17497. doi:10.1371/journal.pone.0017497.

HAWKINS, C. P. 2006. Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. Ecological Applications 16:1277-1294.

HAWKINS, C.P. 2009. Revised Invertebrate RIVPACS Model and O/E Index for Assessing the Biological Condition of Colorado Streams. Colorado Department of Public Health and Environment Water Quality Control Division – Monitoring Unit.

HAWKINS, C. P., Y. CAO, AND B. ROPER. 2010a. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066-1085.

HAWKINS, C. P., R. H. NORRIS, J. GERRITSEN, R. M. HUGHES, S. K. JACKSON, R. K. JOHNSON, AND R. J. STEVENSON. 2000a. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19:541-556.

HAWKINS, C. P., R. H. NORRIS, J. N. HOGUE, AND J. W. FEMINELLA. 2000b. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10:1456-1477.

HAWKINS, C. P., J. R. OLSON, AND R. A. HILL. 2010b. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29:312-343.

HERING, D., C. K. FELD, O. MOOG, AND T. OFENBOECK. 2006. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: Experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566:311-324.

HERLIHY, A. T., S. G. PAULSEN, J. VAN SICKLE, J. L. STODDARD, C. P. HAWKINS, AND L. L. YUAN. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. Journal of the North American Benthological Society 27:860-877.

HOMER, C., J. DEWITZ, J. FRY, M. COAN, N. HOSSAIN, C. LARSON, N. HEROLD, A. MCKERROW, J. N. VANDRIEL, AND J. WICKHAM. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 73:337-341.

KARR, J. R. AND E. W. CHU. 1997. Biological monitoring: essential foundation for ecological risk assessment. Human and Ecological Risk Assessment 3:993-1004.

KOLEFF, P., K. J. GASTON, AND J. J. LENNON. 2003. Measuring beta diversity for presence-absence data. Journal of Animal Ecology 72:367-382.

LEIBOLD, M. A., M. HOLYOAK, N. MOUQUET, P. AMARASEKARE, J. M. CHASE, M. F. HOOPES, R. D. HOLT, J. B. SHURIN, R. LAW, D. TILMAN, M. LOREAU, AND A. GONZALEZ. 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters 7:601-613.

MATTHIESSEN, B., E. MIELKE, AND U. SOMMER. 2010. Dispersal decreases diversity in heterogeneous metacommunities by enhancing regional competition. Ecology 91:2022-2033.

MOSS, D., M. T. FURSE, J. F. WRIGHT, AND P. D. ARMITAGE. 1987. The prediction of the macroinvertebrate fauns of unpolluted running-water sites in Great-Britain using environmental data. Freshwater Biology 17:41-52.

OLSON, J. R., AND C. P. HAWKINS. 2012. Predicting natural base-flow stream water chemistry in the western United States. Water Resources Research 48: W02504, doi:10.1029/2011WR011088.

OSTERMILLER, J. D., AND C. P. HAWKINS. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Journal of the North American Benthological Society 23:363-382.

PFRENDER, M. E., C. P. HAWKINS, M. BAGLEY, G. W. COURTNEY, B. R. CREUTZBURG, J. H. EPLER, S. FEND, L. C. FERRINGTON, P. L. HARTZELL, S. JACKSON, D. P. LARSEN, C. A. LEVESQUE, J. C. MORSE, M. J. PETERSEN, D. RUITER, D. SCHINDEL, AND M. WHITING. 2010. Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. Quarterly Review of Biology 85:319-340.

PONT, D., B. HUGUENY, U. BEIER, D. GOFFAUX, A. MELCHER, R. NOBLE, C. ROGERS, N. ROSET, AND S. SCHMUTZ. 2006. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. Journal of Applied Ecology 43:70-80.

REED, J. C., AND C. A. BUSH. 2001, Generalized Geologic Map of the Conterminous United 782 States, US Geological Survey, Denver, CO.

STODDARD, J. L., A. T. HERLIHY, D. V. PECK, R. M. HUGHES, T. R. WHITTIER, AND E. TARQUINIO. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878-891.

VAN SICKLE, J. 2010. Correlated Metrics Yield Multimetric Indices with Inferior Performance. Transactions of the American Fisheries Society 139:1802-1817.

VAN SICKLE, J., C. P. HAWKINS, D. P. LARSEN, AND A. T. HERLIHY. 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24:178-191.

VAN SICKLE, J., D. D. HUFF, AND C. P. HAWKINS. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwater Biology 51:359-372.

VAN SICKLE, J., D. P. LARSEN, AND C. P. HAWKINS. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. Journal of the North American Benthological Society 26:319-331.

VINSON, M. R., AND C. P. HAWKINS. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society 15:392-399.

WOLOCK, D. M. 1997. STATSGO soil characteristics for the conterminous United States.

WOLOCK, D. M. 2003. Base-flow index grid for the conterminous United States.

WRIGHT, J. F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. Australian Journal of Ecology 20:181-197.

YUAN, L. L., C. P. HAWKINS, AND J. VAN SICKLE. 2008. Effects of regionalization decisions on an O/E index for the US national assessment. Journal of the North American Benthological Society 27:892-905.

Table 2-1. Watershed size classes and land use thresholds used to define reference condition.

Variable	Small Watersheds	Medium Watersheds	Large Watersheds
Area (km^2)	≤ 164	164-3380	> 3380
Count	346	93	61
% Agriculture	0	1	≤ 2
% Urban	0	≤ 1	≤ 2
% Near (3 km) Agriculture	0	≤ 3	≤ 5
% Near (3 km) Urban	0	≤ 3	≤ 5
NDPES discharges (number)	0	NA	NA

Table 2-2. Potential macroinvertebrate metrics used in MMI development.

Name	Description
CHIR-RICH	Chironomid taxa richness
CLING-RICH	Clinger taxa richness
EPHEM-RICH	Ephemeroptera taxa richness
EPT-RICH	Ephemeroptera, Plecoptera, Trichoptera richness
HBIINT1-RICH	Number of taxa with Hilsenhoff family level Biotic Index value = 0, 1, or 2 (intolerant)
HILLEVEN	Hill evenness
INS-RICH	Insect taxa richness
LLT-RICH	Long-lived taxa richness
NON_INS-RICH	Non-insect taxa richness
%CF	Collector-filterer relative abundance
%CHIR	Chironomid relative abundance
%CLING	Clinger relative abundance
%DOMT5	Relative abundance of the five most dominant taxa
%DOMT	Relative abundance of the dominant taxon
%EPHEM	Ephemeroptera relative abundance
%EPT	Ephemeroptera, Plecoptera, Trichoptera relative abundance
%HBIINT1	Relative abundance of taxa with Hilsenhoff family level Biotic Index value = 0, 1, or 2 (intolerant)
%HYDRA	Hydra relative abundance
%PLEC	Plecoptera relative abundance
PLEC-RICH	Plecoptera richness
PREY-RICH	Prey taxa richness
PRED-RICH	Predator taxa richness
RARE1-RICH	number of rare taxa ≤ 1% of total abundance
RARE10-RICH	number of rare taxa ≤ 10% of total abundance
RARE20-RICH	number of rare taxa ≤ 20% of total abundance
RARE5-RICH	number of rare taxa ≤ 5% of total abundance
TOT-RICH	Total taxa Richness
SCRP-RICH	Scraper taxa richness
SHDIVER	Shannon diversity
SHRED-RICH	Shredder taxa richness
SIDIVER	Simpson diversity

Table 2-3. RF models for metrics included in modeled and unadjusted MMI. RF % Var = % variation of reference site metric values explained in RF model. Predictors are arranged by variable importance with decreasing importance from left to right.

Metric	RF % Var	Predictors
SHDIVER	<10.0	
INS-RICH	20.1	ELVmax_WS, WSA, Tmax_WS, PrdCond, Pmin_WS
%EPHEM	21.3	WDmax_WS, Tmax_WS, ELVmax_WS, HYDR_WS, BFI_WS, Pmax_WS
%HYDRA	22.9	Tmax_PT, WSA, Pmax_WS, ELVmin_WS
%CF	23.7	Pmax_PT, ELVmin_WS, Pmax_WS, WS Slope, Tmax_WS, WSA, HYDR_WS
LLT-RICH	31.0	WS Slope, ELVmin_WS, Tmax_WS, PrdCond, Tmax_WS, WSA, HYDR_WS
EPHEM-RICH	32.7	WDmax_WS, WSA, Pmax_WS, PrdCond
NON_INS-RICH	33.9	WDmax_WS, Tmax_WS, HYDR_WS, Pmax_WS, PrdCond, ELVcv_PT
%PLEC	36.8	Pmin_WS, BFI_WS, Pmax_WS, Tmax_WS, WSA
CLING-RICH	39.9	WDmax_WS, WSA, BFI_WS, Pmax_WS

Table 2-4. Metric loadings on varimax rotated principal component axes and t-statistics for tests between reference and most degraded site index score means for modeled (left) and null (right) MMIs. Values highlighted in grey are significantly correlated with one another. Negative t-values imply that a metric increases with degradation, whereas positive t-values imply that a metric decreases with degradation.

		Modeled MMI									Null MMI							
		Axis Loadings									Axis Loadings							
	Metric	1	2	3	4	5	6	7	t	P < 0.05	1	2	3	4	5	6	t	P < 0.05
Richness	INS-RICH	0.83	0.00	0.35	-0.02	0.12	0.10	0.31	10.1	*	0.87	0.33	0.32	0.00	-0.04	-0.03	11.7	*
	RARE5-RICH	0.91	-0.04	0.19	0.06	0.04	-0.17	0.22	9.3	*	0.92	0.22	0.25	0.08	0.04	0.12	10.3	*
	RARE20-RICH	0.88	-0.04	0.33	0.07	0.04	-0.15	0.21	9.2	*	0.89	0.22	0.36	0.08	0.04	0.11	10.5	*
	TOT-RICH	0.88	-0.04	0.34	0.07	0.03	-0.16	0.22	9.1	*	0.89	0.22	0.35	0.07	0.04	0.12	10.3	*
	RARE10-RICH	0.88	-0.03	0.32	0.06	0.02	-0.17	0.22	9.1	*	0.90	0.23	0.32	0.08	0.04	0.12	10.4	*
	RARE1-RICH	0.83	-0.04	-0.03	0.01	-0.01	-0.06	0.14	8.9	*	0.91	0.14	0.06	0.07	0.02	0.06	9.3	*
	PREY-RICH	0.84	-0.13	0.31	0.00	0.04	0.08	0.18	8.5	*	0.86	0.26	0.32	0.02	0.03	-0.02	10.5	*
	PRED-RICH	0.54	0.13	0.25	0.16	0.01	-0.52	0.26	7.8	*	0.67	0.14	0.27	0.15	0.06	0.33	7.7	*
	CHIR-RICH	0.65	-0.45	0.28	0.17	-0.03	0.13	-0.28	4.8	*	0.74	-0.14	0.28	0.15	0.30	-0.38	6.2	*
Composition	%EPT	-0.08	0.77	0.06	0.11	0.33	0.33	0.17	2.3	*	-0.12	0.23	0.04	-0.24	-0.84	-0.03	1.8	NS
	%CHIR	0.13	-0.79	0.08	0.04	-0.13	0.14	-0.12	1.5	NS	0.40	-0.14	0.15	0.19	0.49	-0.52	3.3	*
	%EPHEM	0.00	0.59	0.03	0.35	-0.16	0.46	0.13	1.1	NS	-0.14	-0.04	-0.04	0.14	-0.91	-0.04	-0.4	NS
	%CLING	-0.05	0.68	0.22	-0.38	-0.07	0.01	0.12	0.6	NS	-0.07	0.28	0.03	-0.75	-0.17	0.20	1.8	NS
Dominance/Diversity	SHDIVER	0.46	-0.02	0.85	0.03	0.01	-0.09	0.17	9.8	*	0.58	0.19	0.78	0.08	0.00	0.07	9.8	*
	HILLEVEN	0.20	0.04	0.96	-0.01	0.01	0.00	0.06	6.1	*	0.25	0.14	0.93	0.09	-0.02	0.03	6.1	*
	%DOMT	-0.18	-0.06	-0.91	0.03	-0.06	0.00	-0.09	-6.9	*	-0.26	-0.11	-0.91	-0.03	0.01	-0.03	-7.0	*
	SIDIVER	-0.25	-0.06	-0.93	0.02	-0.03	-0.01	-0.07	-7.4	*	-0.30	-0.13	-0.91	-0.07	0.06	-0.05	-7.4	*
	%DOMT5	-0.45	0.07	-0.77	-0.06	-0.05	0.16	-0.20	-8.5	*	-0.57	-0.22	-0.71	-0.08	-0.05	-0.09	-8.5	*
Composition	%CF	-0.05	-0.17	0.02	-0.73	-0.08	0.11	-0.04	1.0	NS	-0.05	-0.18	-0.10	-0.82	0.14	-0.01	-1.2	NS
	%HYDRA	-0.22	0.27	-0.09	-0.66	-0.02	0.07	0.00	1.0	NS	-0.10	-0.15	-0.12	-0.78	-0.15	-0.05	-2.0	*
Tolerance	%PLEC	0.10	0.04	-0.04	0.00	0.84	0.05	0.06	2.6	*	0.07	0.76	0.16	0.10	0.02	-0.13	4.7	*
	%HBIINT1	-0.03	0.37	0.26	0.14	0.58	0.00	0.25	2.4	*	-0.02	0.65	0.23	0.06	-0.32	0.14	5.5	*
Functional Richness	NON_INS-RICH	0.30	-0.10	0.06	0.26	-0.18	-0.72	-0.10	1.9	NS	0.37	-0.31	0.21	0.28	0.27	0.50	0.4	NS
	CLING-RICH	0.31	0.30	0.17	-0.35	0.06	0.09	0.62	8.5	*	0.47	0.56	0.15	-0.27	-0.31	0.26	9.7	*
	EPT-RICH	0.41	0.25	0.14	-0.02	0.09	0.18	0.76	8.4	*	0.54	0.65	0.15	-0.13	-0.40	0.10	9.3	*
	HBIINT1-RICH	0.25	0.20	0.15	0.06	0.06	0.02	0.86	6.7	*	0.45	0.79	0.08	-0.03	-0.22	0.13	10.5	*
	SCRP-RICH	0.32	0.33	0.02	-0.03	-0.20	-0.01	0.49	6.3	*	0.56	0.40	0.09	-0.14	-0.25	0.09	7.8	*
	EPHEM-RICH	0.37	0.25	0.02	0.16	-0.22	0.42	0.53	5.3	*	0.44	0.27	0.10	-0.11	-0.64	-0.06	5.6	*
	PLEC-RICH	0.05	-0.07	0.10	0.07	0.23	-0.13	0.81	5.3	*	0.36	0.81	0.06	0.02	-0.02	0.11	8.6	*
	SHRED-RICH	0.29	-0.30	0.17	-0.12	0.27	0.06	0.37	4.1	*	0.43	0.62	0.11	0.13	0.09	0.08	7.5	*
	LLT-RICH	0.20	0.07	0.12	-0.38	0.20	-0.28	0.37	3.9	*	0.28	0.21	0.14	-0.12	0.08	0.66	5.0	*

Table 2-5. Index performance measures. RF % Var = % variance of reference site scores associated with natural gradients.

Index	Reference samples (n = 165)			Test samples (n = 416)		Most degraded samples (n = 60)	
	Mean	Precision: CV	Bias: RF % Var	Mean	Sensitivity: % non-ref	Mean	Responsiveness: t-value
O/E-0	1.00	0.21	0.1	0.89	23.3	0.57	12.2
O/E-0-null	1.00	0.23	11.7	0.90	22.6	0.57	12.0
O/E-5	1.06	0.25	0.0	0.94	19.5	0.64	10.5
O/E-5-null	1.00	0.32	7.7	0.87	18.3	0.55	9.3
Modeled MMI*	1.00	0.11	0.0	0.90	34.9	0.64	16.2
Null MMI*	1.00	0.21	22.3	0.89	19.5	0.67	10.7
CO-O/E-0	1.02	0.19					
CO-O/E-5	1.00	0.13					

* MMI scores standardized by dividing by reference score mean.

Table 2-6. Regression models between sample index scores and sample evenness and the change in index precision (CV) after adjustment for evenness.

Index	r^2	slope	Original CV	Adjusted CV	p-value
NV O/E-0	0.33	1.50	0.21	0.18	<0.0001
NV O/E-5	0.19	1.20	0.25	0.24	<0.0001
NV MMI	0.42	1.20	0.11	0.08	<0.0001
CO O/E-0	0.08	0.63	0.18	0.17	0.01
CO O/E-5	0.02	0.28	0.13	0.12	0.11

Table 2-7. Confusion matrices showing differences in assessments of biological condition from raw versus evenness adjusted index scores at NV test sites.

NV O/E-0		Adjusted	
		Degraded	Reference
Raw	Degraded	60	37
	Reference	14	305

NV O/E-5		Adjusted	
		Degraded	Reference
Raw	Degraded	68	13
	Reference	9	326

NV MMI		Adjusted	
		Degraded	Reference
Raw	Degraded	128	17
	Reference	32	239

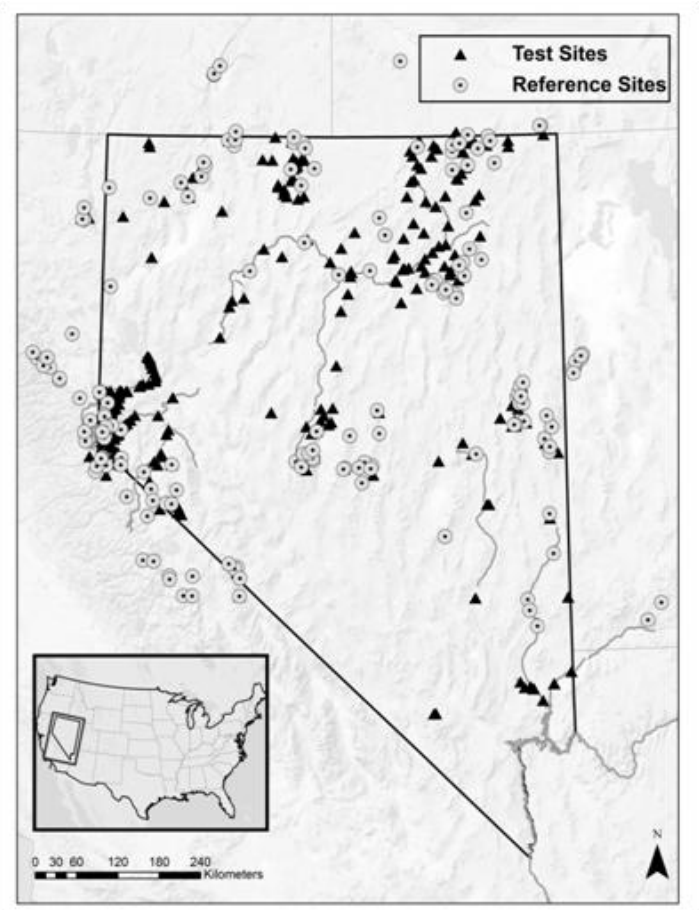


Fig.2-1. Benthic invertebrate sampling locations. Some sites in near proximity to NV were included for index development and evaluations.

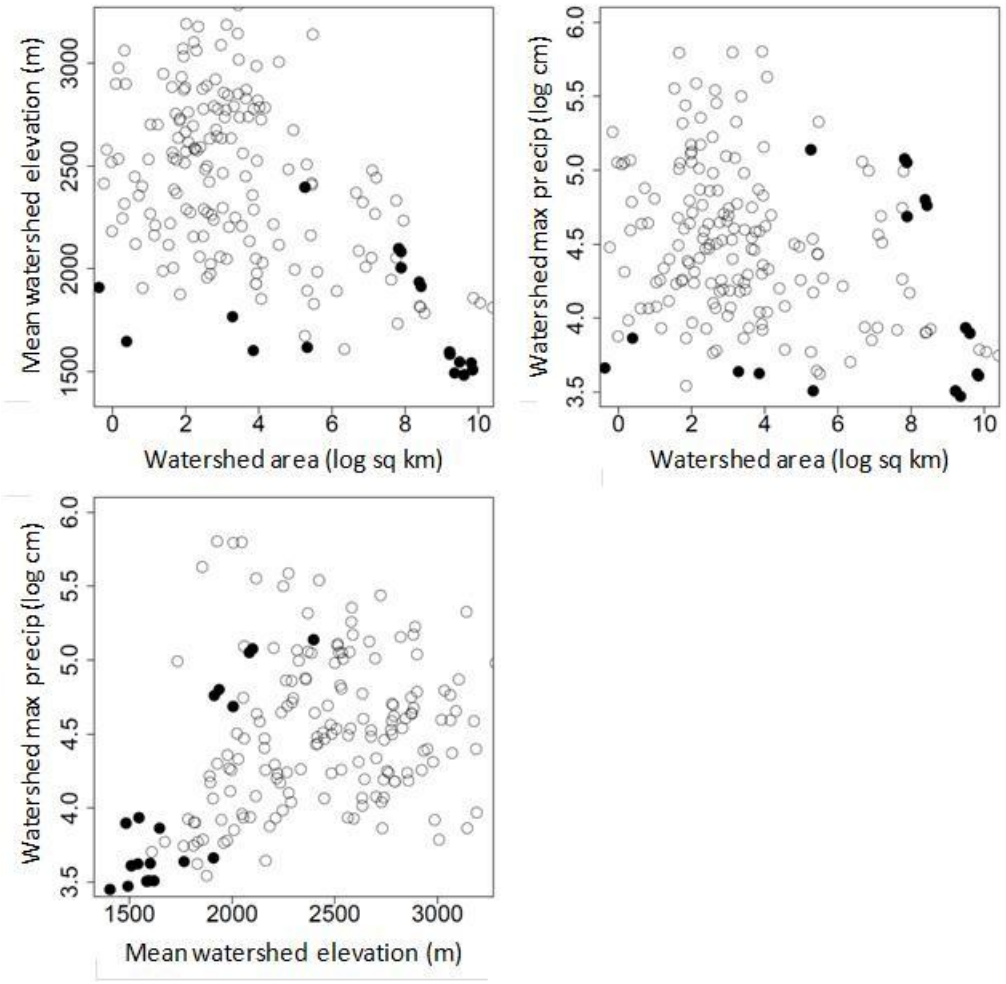


Fig.2-2. Bivariate plots of reference (open circles) and flagged test sites (black circles) along environmental gradients used to detect deviations from model experience in test data.

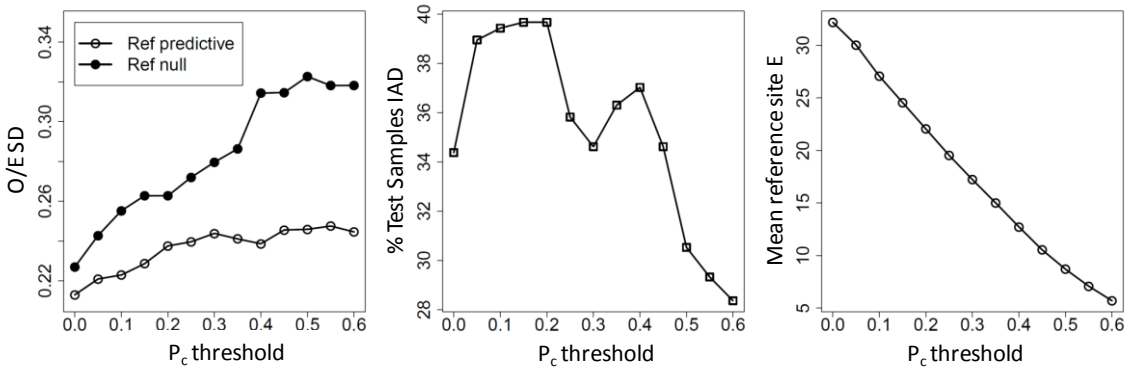


Fig.2-3. Variation in O/E index precision, sensitivity in inferring impairment (IAD = inferred as degraded), and mean expected richness of reference sites with P_c thresholds.

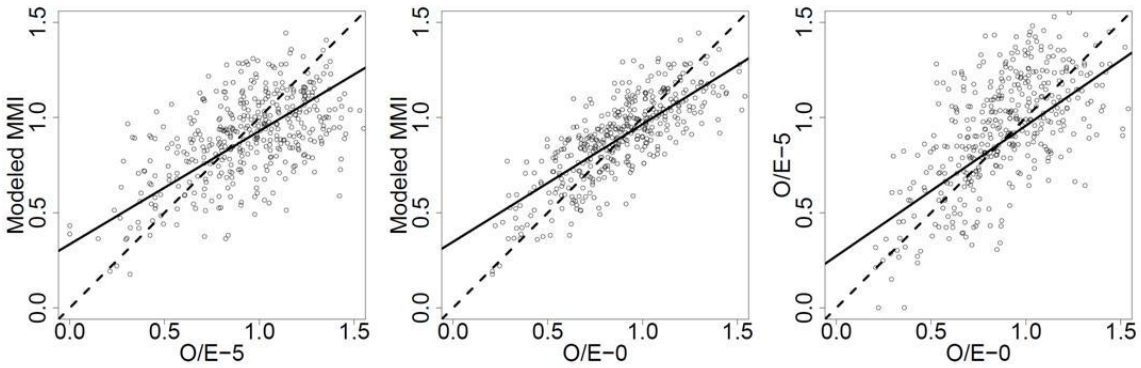


Fig.2-4. Bivariate plots of index scores at test sites. Plots shown with regression lines (solid) and 1:1 lines (dashed) for reference.

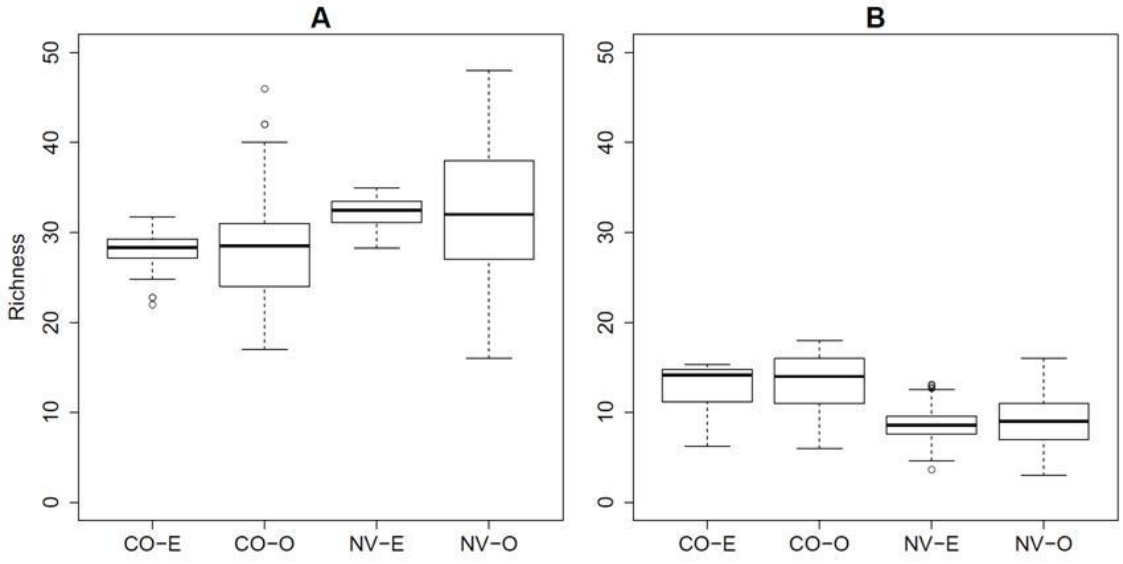


Fig.2-5. Box plots of expected (E) and observed (O) richness for NV and CO samples with Pc > 0 (A) and Pc ≥ 0.5 (B).

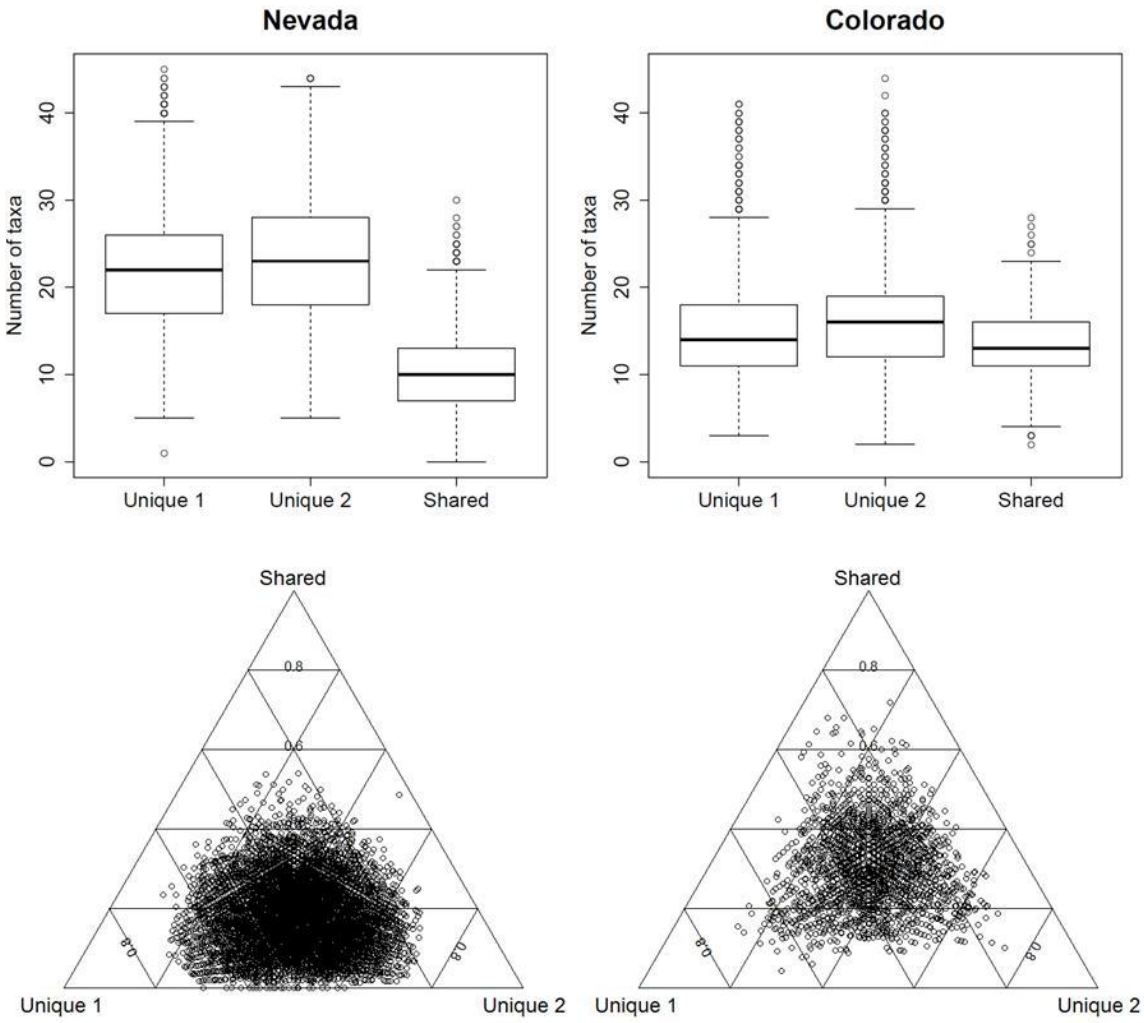


Fig.2-6. Box plots and ternary plots of shared and unique numbers of taxa for all pair-wise comparisons between reference samples from the NV (left) and CO (right).

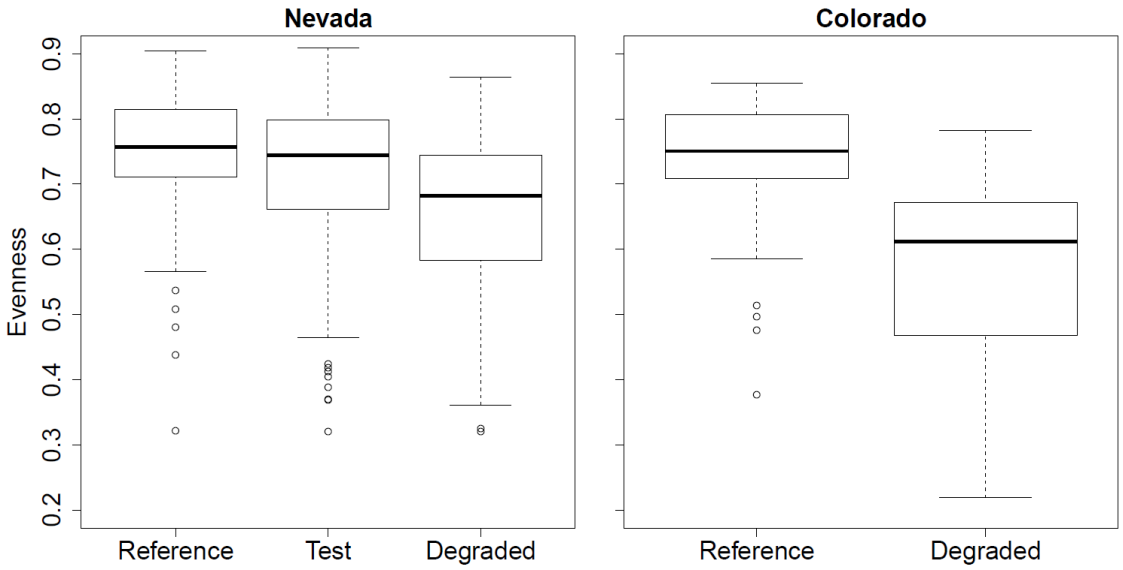


Fig.2-7. Box plots of NV (left) and CO (right) sample evenness by site type.

CHAPTER 3

LINKING LAND USE, INSTREAM STRESSORS, AND BIOLOGICAL CONDITION OF STREAMS TO INFER LIKELY CAUSES OF IMPAIRMENT[2]

Abstract

Bioassessments are important tools for identifying biological impairment in stream ecosystems, but the causes of biological degradation must be identified when developing effective restoration or remediation strategies. Identifying the specific stressors responsible for biological degradation can be challenging because many types of stress represent deviation from natural levels of a physical or chemical attribute rather than the simple addition of a contaminant of anthropogenic origin. Because many natural physical and chemical attributes can vary markedly from stream to stream, stress must be interpreted in the context of site-specific expectations for those attributes. We quantified linkages between biological condition, in-stream stressors, and land use to identify likely causes of biological degradation and their sources in streams of Nevada, USA. We defined potential stress as the site-specific difference between an observed (O) level of a physical or chemical attribute and that expected (E) to occur under natural conditions. We considered four categories of potential stressors: dissolved metals, total dissolved solids, nutrients, and stream temperature. We then used two sets of Random Forest models to quantify relationships between (1) biological condition and O-E_{stress} and (2) measured stressor values and anthropogenic land uses. Based on multiple lines of evidence, we concluded that total dissolved solids (as measured by electrical conductivity) and metal contamination were the stressors most strongly associated with biological degradation in Nevada streams. The most likely sources of excessive levels of these stressors were agriculture,

[2] Coauthor: Charles P. Hawkins

urban development, and mining. Although excessive conductivity has been linked to biological degradation in other streams, conductivity is generally not regulated by current water quality standards. Sufficient evidence now exists to justify developing conductivity criteria for streams. Our results emphasize the importance of interpreting stress in the context of natural variation in stressor levels associated with both point- and non-point sources of pollution. For many pollutants, this approach will require the development of models capable of predicting natural variation in physical and chemical reference conditions.

Introduction

The physiochemical properties of stream water are an important determinant of invertebrate assemblage composition. When human activities cause these properties to exceed their natural range of variation, they can be considered stressors (Townsend et al. 2008) that may alter and degrade ecological communities. Many types of stressors can adversely affect aquatic macroinvertebrate assemblages (e.g. Clements et al. 2000, Van Sickle and Paulsen 2008, Hentges and Stewart 2010, Nicola et al. 2010, Chinnayakanahalli et al. 2011). However, the pathways by which they enter and alter aquatic ecosystems are not always well understood. Anthropogenic activities that can alter physiochemical properties of stream water include urban development (e.g. Paul and Meyer 2001, Walsh et al. 2005), agriculture (e.g. Collins and Jenkins 1996, Richards et al. 1996, Johnson et al. 1997), impoundment (e.g. Preece and Jones 2002), agriculture (e.g. Collins and Jenkins 1996, Johnson et al. 1997, Matthaei et al. 2010), and mineral extraction (e.g. Smolders et al. 2003, Clements 2004). These land uses and the resulting changes in streams are all potential sources of biological degradation.

To conserve and restore the biological condition of streams, we need to be able to identify the causes of biological degradation. Bioassessments are important tools for measuring

biological condition and detecting degradation of stream ecosystems, but alone do not identify the important stressors associated with degradation or the sources of those stressors. Bioassessments are commonly based on a reference site approach in which aspects of the assemblage observed at a site are compared to those that are expected in the absence of human impacts (Hawkins et al. 2010b). Observed biological degradation is assumed to be caused by one or more stressors, but relationships between biological condition and stressors have not been thoroughly assessed (Sloane and Norris 2003). Furthermore, the responsiveness of bioassessment indices to specific stressors is often unknown (Freund and Petty 2007). Models that link bioassessment indices, stressors, and land uses could help verify the sensitivity of bioassessment indices to stressor gradients, quantify the effects of stressors on biological communities, and identify the relative importance of different stressors and their sources.

Although many stressors are known to cause stream degradation, diagnosing the stressors responsible for degradation at specific sites is still problematic. Several frameworks have been suggested to identify the causes of stream degradation (e.g. Suter et al. 2010, Allan et al. 2011, Norris et al. 2012). Because cause and effect can rarely be inferred from single studies (Norris et al. 2012), many of these methods are based on a weight of evidence approach to determine the most likely causes of impairment (Suter et al. 2010). Inferring the cause of degradation requires the identification of a causal mechanism and a consistent and strong association between the hypothesized cause and effect (Norris et al. 2012). Plausible causal mechanisms have been identified for many stressors, but accurately quantifying associations between hypothesized causes and effects can be difficult. This process is often complicated by the presence of natural gradients and the complex effects of co-occurring stressors (Allan et al. 2011).

To accurately attribute degradation of biological condition to stressors, the effects of anthropogenic degradation must be separated from natural variation in both the biota and physicochemical properties of streams. Biota vary along both natural and anthropogenic gradients, and failure to account for natural biotic variation leads to a confounding of natural and anthropogenic effects and increased type I and type II errors of inference (Cao et al. 2007, Hawkins et al 2010a). Bioassessment indices that account for natural biotic variability among sites can be used to isolate the effects of anthropogenic degradation and more confidently attribute variation in index scores to anthropogenic factors. Some stressors may also vary along both natural and anthropogenic gradients, and this natural variability must be accounted for to accurately attribute the presence of stressors to human land uses. However, until recently, there have been few approaches for accounting for natural variability in the spatial distributions of stressors. Similar to the reference condition approach used in bioassessments, models that estimate expected natural physiochemical properties of individual streams could be used to better account for natural variability in stressor distributions. Potential stressors such as stream temperature, electrical conductivity (EC), and nutrients vary along natural gradients, and the degree of alteration that has occurred at a site can be obscured by these strong natural gradients. Currently, models are available to estimate expected site-specific stream EC (Olson and Hawkins 2012a), total nitrogen and phosphorus (Olson and Hawkins 2012b), and summer water temperature (Hill et al. 2012) for our study area, and these models should be useful in accounting for natural variability in these potential stressors.

Accurately modeling complex relationships between stressors and biological condition in multi-stressor environments can also be problematic (Allan 2004, Townsend et al. 2008). Stressors may have interactive effects and can elicit non-linear responses in biota (Townsend et

al. 2008). However, understanding the real-world effects of stressors and identifying meaningful stressor thresholds require the analysis of multi-stressor data (Townsend et al. 2008). Modeling techniques that are robust in detecting interactions among predictors and non-linear responses should improve our ability to quantify complex relationships in multi-stressor environments.

Our main objective in this study was to identify the most likely causes of biological impairment in streams flowing through the state of Nevada (NV), USA. To do this we: 1) quantified individual and cumulative effects of stressors on biological condition, 2) assessed the interactive effects of co-occurring stressors on biota, and 3) identified general sources of different stressors.

Methods

We considered four types of potential stressors to stream biota: total dissolved solids as measured by EC, nutrient enrichment, thermal alteration, and trace metal contamination. To identify important stressors and their sources, we developed two sets of models: one set to link measures of biological condition to in-stream stressors and one set to link spatial distributions of stressors to gradients of land uses (Fig. 3-1). Together these models link watershed-scale land uses to reach-scale stressors, and reach-scale stressors to biological condition of sample sites.

Data and indices

We used two benthic invertebrate based bioassessment indices previously developed for NV streams to characterize biological condition of stream sites (Vander Laan and Hawkins 2012). One index is a multimetric index (MMI) consisting of seven metrics: insect richness, Ephemeroptera relative abundance, Shannon diversity, collector-filterer relative abundance, Plecoptera relative abundance, non-insect richness, and clinger richness. The other index is an

observed to expected (O/E) ratio of community completeness. These indices were developed from samples collected at 165 reference sites throughout NV and surrounding areas (Fig. 3-2). Samples were collected by the NV Department of Environmental Protection (NDEP), the USEPA, and Utah State University. Both indices were based on a modeling approach that accounts for variation in either metric values or taxonomic composition associated with natural environmental gradients. Reference site scores for both of these indices are centered on one, and scores below one indicate anthropogenic degradation.

In addition to the 165 reference samples used in index development, 401 invertebrate samples from test sites throughout NV with corresponding stressor information were available from the NDEP dataset. We assessed biological condition for all test sites by calculating O/E and MMI scores. Water samples were collected by the NDEP during summer low-flow conditions. To characterize chronic stress at test sites, we calculated mean temperature, EC, dissolved metal concentrations, and nutrient concentrations. Means were derived from multiple samples collected at each site between the years 2000 to 2010 (average number of samples per site = 8). EC was measured as µSiemens/cm (µS/cm), and nutrients were measured as concentrations (µg/L) of total nitrogen and total phosphorus. Metal concentrations were measured as dissolved metals (µg/L). We used dissolved metals because this measure is generally considered to be most representative of the biologically available portion of metals in aquatic systems (Reiley 2007). Water chemistry data was also available from the NDEP for 68 of the reference samples used for index development, and we included these samples in the models relating land uses, stressors, and biological condition to ensure that the full ranges of land uses and stressor gradients were represented.

Modeling approach

We developed Random Forest (RF) models to identify and quantify relationships between land uses, stressors, and biological condition. RF models combine predictions from numerous regression or classification trees to produce robust and accurate predictions (Cutler et al. 2007). RF models can be used for both classification and regression, and have significant advantages over other statistical methods such as the ability to model complex interactions among predictors and resistance to model over-fitting (Cutler et al. 2007). Individual relationships between predictors and the response variable are assessed with variable importance measures and partial dependence plots. Variable importance in an RF model is measured as the percent increase in mean squared error of the model when the variable is randomly permuted. Partial dependence plots characterize the effect of an individual predictor on the response by plotting predicted values across the gradient of a predictor while accounting for the effects of all other predictors (Hastie et al. 2001). Because RF can model complex interactions between predictors, it is well suited for assessing individual, cumulative, and interactive effects of stressors on stream invertebrate assemblages. Bivariate partial dependence plots, which characterize the joint effects of two predictors on the response, can be used to assess interactive effects.

Quantifying relationships between biological condition and stressors

To quantify individual relationships between biological condition and individual stressors, we built RF regression models with index scores as responses and stressors as predictors, hereafter biota-stressor models. We developed two biota-stressor models: one with O/E scores as a response and the other with MMI scores as the response. We included all

measured stressors in these models so we could examine both the effects of individual stressors and the potential interactive effects among stressors on biological condition. Failure to account for interactions among correlated stressors could lead to biased assessments of the biological response to individual stressors. We then used univariate and bivariate partial dependence plots of index scores on stressors to assess the associations between biological index values and different stressors and identify meaningful thresholds associated with biological degradation.

For many potential stressors, measured values include contributions from both natural sources and anthropogenic activities. Therefore, the use of measured values in biota-stressor and stressor-land use models could lead to inaccurate assessments of the effects of land uses on stressors and stressors on biota. Of the stressors we examined, TDS, nutrients, and temperature can exhibit marked natural variation among streams. We therefore used existing models to estimate expected site-specific reference conditions for TDS (Olson and Hawkins 2012a), total nitrogen and phosphorus (Olson and Hawkins 2012b), and stream temperature (Hill et al., 2012). We then subtracted expected values from observed values and used these O-E estimates as measures of physiochemical alteration in the biota-stressor and stressor-land use models. Because we lacked a practical way of assessing natural variation in metal concentrations among sites, we assumed background concentrations were zero, i.e., O-E = O.

To quantify the cumulative effects of stressors on biological condition, we developed RF models to predict biological condition classes (i.e. reference or impaired) from stressor information. We classified sites as being in either reference or impaired biological condition based on a one-sided, non-central equivalence test that compared the score of an assessed site with the distribution of reference site scores (Vander Laan and Hawkins 2012). We considered a site to be in reference condition if the index score was inferred to be above the fifth percentile

of the distribution of reference site scores with 95% confidence. We considered sites with index scores below this threshold to be impaired. We then developed RF models to predict biological state (reference or impaired) as a function of the mix of stressors measured at each site.

Relating stressors to land uses

We developed several RF models to relate biologically important stressors (i.e. those used in biota-stressor models) to land uses within watersheds. We included both land use and natural environmental characteristics as predictors of measured stressor levels to determine the relative importance of human versus natural factors in driving variability in stressors among streams. We used a geographic information system (GIS) to characterize both land use and the natural environmental characteristics of each watershed (Appendix A). We obtained land cover information from the National Land Cover Database (NLCD) produced by the Multi-Resolution Land Characteristics Consortium (Homer et al. 2007) to calculate the percent of area in each watershed that was classified as agricultural or urban land use and the percent of the watershed within 3 km upstream from the sample site that was classified as agricultural or urban land use. We obtained mining data from the USGS mineral resources data system (http://mrdata.usgs.gov) and characterized mining activity as the density of known mine sites in each watershed. We characterized dams within the watershed as the total dam volume and the volume of the largest dam within each watershed (both standardized by dividing by watershed area) as defined by the National Inventory of Dams (USACE, 2006). We extracted thirty-year average air temperature and precipitation information from grids produced by the PRISM climate group (Daly et al. 2008). We used geologic (Reed and Bush 2001) and soil (Wolock 1997) data to define geology and soil properties of each watershed. We characterized watershed topography as watershed size; the mean, minimum, and maximum watershed elevation; and

watershed slope calculated as the change in watershed elevation (maximum elevation minus minimum elevation) divided by the maximum flow length. We also used two measures to characterize hydrologic regime: (1) mean base flow index (Wolock 2003) and (2) a measure of hydrologic stability calculated as the minimum mean monthly discharge divided by the maximum mean monthly discharge interpolated from USGS gauging stations.

We developed models iteratively by including all potential predictors first, and then removing predictors that had low predictive power, were highly correlated with other predictors, or lacked an interpretable relationship with the response. For nutrients, TDS, and temperature, we included model derived expected values as predictors and excluded all other natural predictors. For metals, potential predictors included all measures of land uses and all measures of natural environmental characteristics (Appendix A). We retained anthropogenic factors with interpretable relationships with the response variables, even if they were relatively weak predictors so that we could assess associations between land uses and stressors. We used the percent variation in each stressor explained by the models, variable importance measures, and partial dependence plots to assess the performance of models and interpret relationships between land use and stressors.

Results

Stressor levels

Stressor levels varied considerably among sites (Table 3-1), although they were relatively low at most sites. Arsenic was the most variable potential stressor followed by total P and EC.

Relationships between biological condition and stressors

O/E scores were negatively associated (R^2=0.22) with O-E EC, copper, zinc, and O-E stream temperature (Fig. 3-3). O-E EC and arsenic were the most important predictors of O/E_{taxa}, and produced the strongest negative responses (Fig. 3-4). O/E_{taxa} declined steadily with increases in EC above expected natural conditions and with increasing arsenic. Decreases in O/E_{taxa} were also weakly associated with increases in copper and zinc and with cooling stream temperature.

MMI scores were also negatively associated (R^2=0.15) with a suite of stressors including arsenic, O-E EC, O-E temperature, and O-E nutrients (Fig. 3-3). Arsenic, O-E temperature, and O-E EC were the most important predictors of MMI scores (Fig. 3-4). Decreases in MMI scores were most strongly associated with arsenic (Fig. 3-3). Decreases in MMI were also associated with increases in O-E EC, O-E total N, O-E total P, and decreases in stream temperature. Unlike O/E_{taxa}, the MMI was generally insensitive to variation in zinc and copper (Fig. 3-3).

Interactive effects of stressors

Bivariate partial dependence plots indicated that co-occurring stressors generally had additive effects on biota (Fig. 3-5). For example, the joint effects of elevated EC and zinc on O/E_{taxa} were greater than the individual effects of elevated levels of either stressor. For O/E_{taxa}, we observed similar patterns for all combinations of stressors. In the MMI model, additive effects were also apparent with combinations of O-E EC and arsenic, O-E EC and O-E temperature, and O-E total P and O-E total N. However, we also observed small mitigating effects of increased nutrients on arsenic (Fig. 3-5).

Inferring biological condition class from stressors

Stressors were generally good predictors of biological condition class. RF models predicted O/E_{taxa} biological condition classes with about 79% accuracy and MMI condition classes with about 70% accuracy. Although overall accuracy of predictions of biological condition was generally high, the prediction error for degraded sites was much higher than for reference sites (Table 3-2).

Relating stressors to land uses

In-channel stressor levels were associated with land use gradients and environmental characteristics (Appendix B). The RF models accounted for between 32% and 93% of spatial variation in stressors (Table 3-3). Altered EC was the most predictable stressor, and arsenic was the least predictable stressor. Land uses were important predictors of all stressors except arsenic (Table 3-3). For all stressors, except O-E total P, the percent of the entire watershed classified as urban or agriculture was a better predictor of stressor levels than the percent of the watershed within 3 km upstream from the sample site. Metals such as copper and zinc tended to be associated with urban development and mining within watersheds. Altered EC and nutrient levels were positively associated with agriculture and urban development. Stream temperature was also positively related to agriculture, but negatively associated with urbanization.

Discussion

Restoring the biological integrity of degraded streams requires that we identify the stressors that are causing degradation and the sources of those stressors. Although bioassessments are useful tools for identifying biological degradation, they do not identify the

causes of this degradation. We used a modeling approach to quantify relationships between biological index scores, potential stressors, and measures of watershed alteration. In doing so, we addressed two major challenges for quantifying relationships between biota, stressors, and land uses: 1) separating anthropogenic effects on stream ecosystems from natural variation in both stressors and biota, and 2) assessing the relative, cumulative, and interactive effects of co-occurring stressors on biotic condition. To identify the most likely causes of biological degradation, we built on weight of evidence approaches developed by others (e.g. Suter et al. 2010, Allan et al. 2011, Norris et al. 2012). In a weight of evidence approach, confidence in inferences increases when observed relationships between stressors and biota are consistent with established causal mechanisms.

Index responsiveness to stressors

As we observed, different biological indices may be differentially responsive to the same stressors. These differences in responsiveness have implications for interpreting the biological effects of stressors on aquatic ecosystems. The higher responsiveness of the O/E_{taxa} index than the MMI to stressors may have occurred for at least two reasons. First, differences in how indices are calibrated may affect their responsiveness (Hawkins et al. 2010a). For example, O/E_{taxa} indices are calibrated with only reference data, whereas MMIs are calibrated with both reference sites and pre-defined degraded sites. Although calibrating an index with data from disturbed sites should generally lead to high responsiveness, MMIs may show dampened response to high levels of stress or to novel stressors if the degraded sites used in calibration do not adequately characterize the complete stressor gradient within a region (Hawkins et al. 2010a). Second, O/E_{taxa} and an MMI measure somewhat different biological properties of the same assemblage, The responsiveness of an MMI may be dampened by the aggregation of

information from individual taxa into composite metrics that describe community-level attributes (trophic structure, diversity, etc.). Because sensitivity to stressors varies among taxa that contribute information to a metric, the overall responsiveness of a metric should be some average function of responses of those specific taxa that contribute to a metric. Furthermore, an MMI as a whole is comprised of individual metrics that may differ in their response to any given stress. O/E_{taxa} is not prone to these averaging effects, because reductions in O/E occur when individual taxa expected at a site are lost, presumably due to stress and in order of their sensitivities to local stressors.

Accounting for natural variability in both biota and stressors

Many bioassessments and causal analyses are potentially confounded by spatial co-variation of naturally occurring features and human alteration of the environment. For example, more human-associated alteration has occurred in lowland than upland settings. Our use of modeled bioassessment indices (Vander Laan and Hawkins 2012) allowed us to account for the effects of naturally occurring environmental variability on biota and more confidently attribute changes in stream macroinvertebrates to anthropogenic stressors. We were able to use a similar approach to better understand when physicochemical conditions likely exceeded levels expected at individual sites, which allowed us to more accurately describe both biota-stress relationships and stress-land use relationships.

The quantification of stressors can be an especially difficult problem when analyzing field data. For novel stressors such as pesticides, a direct measure of the concentrations observed at each site is a relevant estimate of exposure because natural background concentrations must be zero. In this study, we assumed that background concentrations of trace metals were also zero. This assumption may not have been completely robust, though, because

relatively high levels of metals can result from natural geologic sources (Schmidt et al. 2012). We were not able to account for potential natural sources of trace metals in our study area. However, the levels of metal concentrations that we observed at reference sites were below those considered toxic (http://water.epa.gov/scitech/swguidance/standards/criteria/current/index.cfm) which suggests that natural background metal concentrations may be generally low in the study region. Many other stressors represent human-caused changes in physicochemical conditions that naturally vary among locations, e.g., water chemistry, temperature, sediment, and nutrients. In these cases, potential stress is best measured as deviation from natural conditions. Because we were able to estimate deviation from expected reference condition for EC (Olson and Hawkins 2012a), nutrients (Olson and Hawkins 2012b), and temperature (Hill et al. 2012), we were able to strengthen inferences regarding both the effect alterations in these factors have on biota and the land use activities that are associated with their alteration.

Interactive effects of EC and metals on biota

We expected to see an ameliorating effect of increasing EC on the effects of metal contamination on O/E_{taxa} and MMI values, because the bioavailability, and hence toxicity, of metal ions is reduced by the prevalence of other dissolved ions (Clements et al. 2000, Schmidt et al. 2010). However, our observation of additive effects of increases in O-E EC and metal contamination (Fig. 3-5) suggests that any ameliorating effects of increased EC on metal toxicity were swamped by the differences in metal concentrations that occurred among sites. These results imply that scaling metal toxicity to local water chemistry conditions, as is done with cumulative criterion units (e.g. Clements et al. 2000, Schmidt et al. 2010), may not always be needed. We concluded that the RF models incorporated interactions between stressors and this

provided accurate estimates of each stressor's effect on stream biota in the context of other stressors and naturally occurring factors. However, this interpretation needs to be tested within an experimental framework in which O/E_{taxa}, MMI values, and survival of individual taxa are assessed in response to joint variation in metal concentrations and EC.

Stressor-specific relationships with biota and land uses

Metals – Zinc, copper, and arsenic were all associated with degraded biological condition in our dataset, which is consistent with several other studies that report negative associations between the condition of invertebrate assemblages and metal contamination under either field (e.g. Clements et al. 2000, Cain et al. 2004, Pollard and Yuan 2006) or laboratory settings (e.g. Richardson and Kiffney 2000, Clements et al. 2002, Clements 2004). Furthermore, the causal mechanisms by which metals affect stream invertebrate communities are well established. Chronic metal contamination can increase invertebrate drift (Clements 2004), alter ecosystem functions (Carlisle and Clements 2005), and reduce the fitness of some aquatic invertebrates (e.g. Wicklum and Davies 1996). This weight of evidence strongly supports the inference that metal contamination is a stressor of concern in NV streams. Zinc and copper were both strongly associated with human land uses such as urbanization and mining, which implies that these land uses may result in increased levels of contamination from these metals and subsequently, biological degradation. These relationships are consistent with previous findings (e.g. Paulson 1997, Beasley and Kneale 2002, Macklin et al. 2006, Wong et al. 2006, Xiao and Ji 2007) and indicate that these land uses contribute to metal contamination and biological degradation in NV and other streams.

Arsenic was also associated with degraded biological condition, but it was not associated with land use gradients. It is unclear whether the biological associations with arsenic

that we observed can be attributed to anthropogenic degradation or if they result from naturally occurring arsenic in stream waters (e.g. Wilkie and Hering 1998). Determining if elevated arsenic levels result from natural conditions or human activities will require a detailed analysis of potential natural sources or arsenic, which was beyond the scope of this study. However, the toxic effects of arsenic on stream invertebrates are well established, and the observed associations with biological condition may still be relevant to streams with naturally low levels of arsenic. In general, improved ways to account for background metal concentrations in bioassessments and causal analyses (e.g. Schmidt et al. 2012) would benefit our understanding of how metal contamination and its possible sources may degrade stream ecosystems.

Temperature – The associations we observed between stream cooling and biological degradation (Fig. 3-3) and between cooling and the presence of dams within a watershed (Table 3-3) indicates that the cooling effects of large dams in a watershed can lead to biological degradation. However, these dams also cause a suite of other environmental changes (i.e. flow regime, stream bed characteristics) that may be confounded with decreases in stream temperature and have major effects on invertebrate assemblages (Vinson 2001). Although partial dependence plots did not show a strong dependence of either O/E_{taxa} or MMI on increases in temperature, the two lowest scores we observed for both indices were associated with extremely high stream temperatures (>34 ˚C). However, these samples were from sites influenced by hot-springs, and the apparent degradation we observed is unlikely to be associated with anthropogenic stressors. Bioassessment indices generally use watershed or reach scale environmental characteristics to predict expected invertebrate assemblages, and are therefore unable to account for the effects of such highly localized natural environmental

conditions. The response of these indices to high stream temperatures does indicate that they should be useful for detecting degradation caused by anthropogenic warming where it may occur.

Nutrients – The O/E_{taxa} index and the MMI were unresponsive or only weakly responsive to nutrient enrichment. As opposed to stressors such as metals, temperature, and EC, it is unlikely that nutrient enrichment has a direct effect on stream invertebrate assemblages. Although oxygen depletion resulting from nutrient enrichment can have major effects on invertebrate assemblages, these effects are most common in lakes and estuaries, and are generally uncommon and localized in streams (Allan 2004). Because nutrient enrichment in streams is more likely to have a direct effects on algal assemblages, indices based on algae (e.g. Smith et al. 2007) may be better indicators of eutrophication than indices based on invertebrates. In fact, in naturally nutrient poor streams, modest increases in nutrients may increase macroinvertebrate abundance and richness and potentially compensate for the adverse effects of other stressors (e.g. Fig. 3-5, Hawkins et al. 2000).

Total dissolved solids – The effects of EC on stream biota are less well established than that for many other stressors. However, increased EC was one of the most important predictors of biological degradation in NV streams. We observed biological degradation associated with any increase in EC above expected natural levels (i.e. O-E EC > 0). For our study area, this translates to absolute conductivities around 300 µS/cm or lower. This level of EC is well below levels associated with acute toxicity for most freshwater invertebrates (Blasius and Merritt 2002, Kefford et al. 2003, Benbow and Merritt 2004, Kefford et al. 2005). Our results suggest that assemblage degradation associated with elevated EC may not result from toxicity, but from changes in assemblage composition stemming from changes in osmoregulatory niche space.

Increases in EC improve conditions for taxa that are intolerant of very low ionic strength water, and taxa that are adapted for low EC may be outcompeted and excluded from systems with elevated EC (Olson and Hawkins 2012c). The 300 µS/cm threshold we observed is similar to thresholds observed in Central Appalachian streams (U.S. EPA 2011), and may indicate a major natural threshold between freshwater invertebrate assemblage types. Although the primary driver of variation in EC among streams in our study was natural (Table 3-3), land uses such as agriculture, urbanization, and mining were also important predictors. Similar to our findings, others have also observed increases in stream EC associated with agriculture (e.g. Johnson et al. 1997, Pan et al. 2004), urbanization (e.g. Wang 1997, Hatt et al. 2004), and mining (e.g. Pond et al. 2008, Palmer et al. 2010). The strong association between altered EC and these land uses implies that watershed alteration can lead to ECs that exceed natural ranges and therefore result in degraded biological condition.

Implications for causal analysis of stream degradation

Inferring the causes of biological degradation in streams is a difficult process. However, we cannot effectively protect or restore stream ecosystems without understanding the causes of biological degradation at both local and regional spatial scales. In particular, causal analysis is complicated by two factors: 1) stressors that may vary along both natural and anthropogenic gradients and 2) the spatial co-variation of, and possible interactive effects between, stressors. Models that predict expected natural levels of EC, nutrients, and stream temperature (e.g. Hawkins et al. 2010b, Olson and Hawkins 2012a,b, Hill et al. 2012) can be used to account for some of the natural variability in these potential stressors and thus more accurately quantify levels of stress. However, similar, easily applied models do not yet exist for other important

stressors like sedimentation or acidification, and EC and nutrient models have only been developed for streams of the western United States.

Additional research on how changes in EC affect freshwater communities is especially needed. Although we and others (e.g. Pond et al. 2008, U.S. EPA 2011) have linked changes in EC to changes in invertebrate assemblages, the causal mechanisms by which EC affects invertebrate fitness and survival are not well understood. Further work that establishes the physiological and ecological bases for invertebrate assemblage shifts near 300 µS/cm would greatly strengthen our understanding of this potentially critical stressor and how to best develop EC criteria that are protective of freshwater ecosystems.

Literature Cited

ALLAN, J.D., L.L. YUAN, P. BLACK, T. STOCKTON, P.E. DAVIES, R.H. MAGIEROWSKI, AND S.M. READ. 2011. Investigating the relationships between environmental stressors and the stream condition using Bayesian belief networks. Freshwater Biology doi: 10.1111/j.1365-2427.2011.02683.x.

ALLAN, J. D. 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. Annual Review of Ecology Evolution and Systematics 35:257-284.

BEASLEY, G., AND P. KNEALE. 2002. Reviewing the impact of metals and PAHs on macro invertebrates in urban watercourses. Progress in Physical Geography 26:236-270.

BENBOW, M. E., AND R. W. MERRITT. 2004. Road-salt toxicity of select Michigan wetland macroinvertebrates under different testing conditions. Wetlands 24:68-76.

BLASIUS, B. J., AND R. W. MERRITT. 2002. Field and laboratory investigations on the effects of road salt (NaCl) on stream macroinvertebrate communities. Environmental Pollution 120:219-231.

CAIN, D. J., S. N. LUOMA, AND W. G. WALLACE. 2004. Linking metal bioaccumulation of aquatic insects to their distribution patterns in a mining-impacted river. Environmental Toxicology and Chemistry 23:1463-1473.

CAO, Y., C. P. HAWKINS, J. OLSON, AND M. A. KOSTERMAN. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26:566-585.

CHINNAYAKANAHALLI, K. J., C. P. HAWKINS, D. G. TARBOTON, AND R. A. HILL. 2011. Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. Freshwater Biology 56:1248-1265.

CARLISLE, D. M., AND W. H. CLEMENTS. 2005. Leaf litter breakdown, microbial respiration and shredder production in metal-polluted streams. Freshwater Biology 50:380-390.

CLEMENTS, W. H. 2004. Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. Ecological Applications 14:954-967.

CLEMENTS, W. H., D. M. CARLISLE, L. A. COURTNEY, AND E. A. HARRAHY. 2002. Integrating observational and experimental approaches to demonstrate causation in stream biomonitoring studies. Environmental Toxicology and Chemistry 21:1138-1146.

CLEMENTS, W. H., D. M. CARLISLE, J. M. LAZORCHAK, AND P. C. JOHNSON. 2000. Heavy metals structure benthic communities in Colorado mountain streams. Ecological Applications 10:626-638.

COLLINS, R., AND A. JENKINS. 1996. The impact of agricultural land use on stream chemistry in the Middle Hills of the Himalayas, Nepal. Journal of Hydrology 185:71-86.

CUTLER, D. R., T. C. EDWARDS, K. H. BEARD, A. CUTLER, AND K. T. HESS. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

DALY, F., C., M. HALBLEIB, J. I. SMITH, W. P. GIBSON, M. K. DOGGETT, G. H. TAYLOR, J. CURTIS, AND P. P. PASTERIS. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology 28:2031-2064.

FREUND, J. G., AND J. T. PETTY. 2007. Response of fish and macroinvertebrate bioassessment indices to water chemistry in a mined appalachian watershed. Environmental Management 39:707-720.

HASTIE, T. J., R. J. TIBSHIRANI, AND J. H. FRIEDMAN. 2001. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics. Spring, New York, New York, USA.

HATT, B. E., T. D. FLETCHER, C. J. WALSH, AND S. L. TAYLOR. 2004. The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. Environmental Management 34:112-124.

HAWKINS, C. P., Y. CAO, AND B. ROPER. 2010a. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066-1085.

HAWKINS, C. P., R. H. NORRIS, J. N. HOGUE, AND J. W. FEMINELLA. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10:1456-1477.

HAWKINS, C. P., J. R. OLSON, AND R. A. HILL. 2010b. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29:312-343.

HENTGES, V. A., AND T. W. STEWART. 2010. Macroinvertebrate Assemblages in Iowa Prairie Pothole Wetlands and Relation to Environmental Features. Wetlands 30:501-511.

HILL, R.A., HAWKINS, C.P., AND D.M. CARLISLE. 2012. Predicting thermal reference conditions in USA streams. Freshwater Science (in review).

HOMER, C., J. DEWITZ, J. FRY, M. COAN, N. HOSSAIN, C. LARSON, N. HEROLD, A. MCKERROW, J. N. VANDRIEL, AND J. WICKHAM. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 73:337-341.

JOHNSON, L. B., C. RICHARDS, G. E. HOST, AND J. W. ARTHUR. 1997. Landscape influences on water chemistry in Midwestern stream ecosystems. Freshwater Biology 37:193-208.

KEFFORD, B. J., C. G. PALMER, AND D. NUGEGODA. 2005. Relative salinity tolerance of freshwater macroinvertebrates from the south-east Eastern Cape, South Africa compared with the Barwon Catchment, Victoria, Australia. Marine and Freshwater Research 56:163-171.

KEFFORD, B. J., P. J. PAPAS, AND D. NUGEGODA. 2003. Relative salinity tolerance of macroinvertebrates from the Barwon River, Victoria, Australia. Marine and Freshwater Research 54:755-765.

MACKLIN, M. G., P. A. BREWER, K. A. HUDSON-EDWARDS, G. BIRD, T. J. COULTHARD, I. A. DENNIS, P. J. LECHLER, J. R. MILLER, AND J. N. TURNER. 2006. A geomorphological approach to the management of rivers contaminated by metal mining. Geomorphology 79:423-447.

MATTHAEI, C. D., J. J. PIGGOTT, AND C. R. TOWNSEND. 2010. Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. Journal of Applied Ecology 47:639-649.

NICOLA, G. G., A. ALMODOVAR, AND B. ELVIRA. 2010. Effects of environmental factors and predation on benthic communities in headwater streams. Aquatic Sciences 72:419-429.

NORRIS, R.H., J.A. WEBB, S.J. NICHOLS, M.J. STEWARDSON, AND E.T. HARRISON. 2012. Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshwater Science 31:5-21.

OLSON, J. R., AND C. P. HAWKINS. 2012a. Predicting natural base-flow stream water chemistry in the western United States. Water Resources Research 48.

OLSON, J. R., AND C. P. HAWKINS. 2012b. Developing site-specific nutrient criteria from empirical models. Freshwater Science (in review).

OLSON, J. R., AND C. P. HAWKINS. 2012c. The Influence of Geology and Other Environmental Factors on Stream Water Chemistry and Benthic Invertebrate Assemblages. Functional Ecology (in review).

PALMER, M. A., E. S. BERNHARDT, W. H. SCHLESINGER, K. N. ESHLEMAN, E. FOUFOULA-GEORGIOU, M. S. HENDRYX, A. D. LEMLY, G. E. LIKENS, O. L. LOUCKS, M. E. POWER, P. S. WHITE, AND P. R. WILCOCK. 2010. Mountaintop Mining Consequences. Science 327:148-149.

PAN, Y. D., A. HERLIHY, P. KAUFMANN, J. WIGINGTON, J. VAN SICKLE, AND T. MOSER. 2004. Linkages among land-use, water quality, physical habitat conditions and lotic diatom assemblages: A multi-spatial scale assessment. Hydrobiologia 515:59-73.

PAUL, M. J. AND J. L. MEYER. 2001. Streams in the urban landscape. Annual Review of Ecology and Systematics 32:333-365.

PAULSON, A. J. 1997. The transport and fate of Fe, Mn, Cu, Zn, Cd, Ph and SO4 in a groundwater plume and in downstream surface waters in the Coeur d'Alene Mining District, Idaho, USA. Applied Geochemistry 12:447-464.

POLLARD, A. I., AND L. YUAN. 2006. Community response patterns: evaluating benthic invertebrate composition in metal-polluted streams. Ecological Applications 16:645-655.

POND, G. J., M. E. PASSMORE, F. A. BORSUK, L. REYNOLDS, AND C. J. ROSE. 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. Journal of the North American Benthological Society 27:717-737.

PREECE, R. M., AND H. A. JONES. 2002. The effect of Keepit Dam on the temperature regime of the Namoi River, Australia. River Research and Applications 18:397-414.

REED, J. C., AND C. A. BUSH. 2001., Generalized Geologic Map of the Conterminous United 782 States, US Geological Survey, Denver, CO.

REILEY, M. C. 2007. Science, policy, and trends of metals risk assessment at EPA: how understanding metals bioavailability has changed metals risk assessment at US EPA. Aquatic Toxicology 84:292-298.

RICHARDS, C., L. B. JOHNSON, AND G. E. HOST. 1996. Landscape-scale influences on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Sciences 53:295-311.

RICHARDSON, J. S., AND P. M. KIFFNEY. 2000. Responses of a macroinvertebrate community from a pristine, southern British Columbia, Canada, stream to metals in experimental mesocosms. Environmental Toxicology and Chemistry 19:736-743.

SCHMIDT, T. S., W. H. CLEMENTS, K. A. MITCHELL, S. E. CHURCH, R. B. WANTY, D. L. FEY, P. L. VERPLANCK, AND C. A. SAN JUAN. 2010. Development of a new toxic-unit model for the bioassessment of metals in streams. Environmental Toxicology and Chemistry 29:2432-2442.

SCHMIDT, T. S., W. H. CLEMENTS, R. B. WANTY, P. L. VERPLANCK, S. E. CHURCH, C. A. SAN JUAN, D. L. FEY, B. W. ROCKWELL, E. H. DEWITT, AND T. L. KLEIN. 2012. Geologic processes influence the effects of mining on aquatic ecosystems. Ecological Applications 22:870-879.

SLOANE, P. I. W., AND R. H. NORRIS. 2003. Relationship of AUSRIVAS-based macroinvertebrate predictive model outputs to a metal pollution gradient. Journal of the North American Benthological Society 22:457-471.

SMITH, A. J., R. W. BODE, AND G. S. KLEPPEL. 2007. A nutrient biotic index (NBI) for use with benthic macroinvertebrate communities. Ecological Indicators **7**:371-386.

SMOLDERS, A. J. P., R. A. C. LOCK, G. VAN DER VELDE, R. I. M. HOYOS, AND J. G. M. ROELOFS. 2003. Effects of mining activities on heavy metal concentrations in water, sediment, and macroinvertebrates in different reaches of the Pilcomayo River, South America. Archives of Environmental Contamination and Toxicology 44:314-323.

SUTER, G. W., II, S. B. NORTON, AND S. M. CORMIER. 2010. The science and philosophy of a method for assessing environmental causes. Human and Ecological Risk Assessment 16:19-34.

TOWNSEND, C. R., S. S. UHLMANN, AND C. D. MATTHAEI. 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45:1810-1819.

USACE (UNITED STATES ARMY CORPS OF ENGINEERS). 2006. National Inventory of Dams.

U.S. EPA (ENVIRONMENTAL PROTECTION AGENCY). 2011. A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams. Office of Research and Development, National Center for Environmental Assessment, Washington, DC. EPA/600/R-10/023.

VAN SICKLE, J., AND S. G. PAULSEN. 2008. Assessing the attributable risks, relative risks, and regional extents of aquatic stressors. Journal of the North American Benthological Society 27:920-931.

VANDER LAAN, J.J., AND C.P. HAWKINS. 2012. Assessing biological condition of arid zone streams: Effects of modeling, spatial isolation, and sample evenness. Freshwater Science (in review).

VINSON, M. R. 2001. Long-term dynamics of an invertebrate assemblage downstream from a large dam. Ecological Applications 11:711-730.

WALSH, C. J., A. H. ROY, J. W. FEMINELLA, P. D. COTTINGHAM, P. M. GROFFMAN, AND R. P. MORGAN. 2005. The urban stream syndrome: current knowledge and the search for a cure. Journal of the North American Benthological Society 24:706-723.

WANG, X. H., AND Z. Y. YIN. 1997. Using GIS to assess the relationship between land use and water quality at a watershed level. Environment International 23:103-114.

WICKLUM, D., AND R. W. DAVIES. 1996. The effects of chronic cadmium stress on energy acquisition and allocation in a freshwater benthic invertebrate predator. Aquatic Toxicology 35:237-252.

WILKIE, J. A., AND J. G. HERING. 1998. Rapid oxidation of geothermal arsenic(III) in streamwaters of the eastern Sierra Nevada. Environmental Science & Technology 32:657-662.

WOLOCK, D. M. 1997. STATSGO soil characteristics for the conterminous United States.

WOLOCK, D. M. 2003. Base-flow index grid for the conterminous United States.

WONG, C. S. C., X. D. LI, AND I. THORNTON. 2006. Urban environmental geochemistry of trace metals. Environmental Pollution 142:1-16.

XIAO, H., AND W. JI. 2007. Relating landscape characteristics to non-point source pollution in mine waste-located watersheds using geospatial techniques. Journal of Environmental Management 82:111-119.

Table 3-1. Minimum, maximum, and coefficient of variation of potential stressors across sites.

Stressor	min	max	CV
EC (µS/cm)	22.00	3256.20	1.36
Total P (µg/L)	0.00	2.60	1.97
Total N (µg/L)	0.00	4.20	1.05
Temperature (˚C)	6.16	34.81	NA
Cadmium (µg/L)	1.00	2.00	0.14
Arsenic (µg/L)	1.92	250.67	2.10
Chromium (µg/L)	1.50	4.00	0.10
Copper (µg/L)	1.67	5.67	0.17
Lead (µg/L)	1.00	2.33	0.26
Mercury (µg/L)	0.12	0.46	0.34
Nickel (µg/L)	2.79	16.00	0.22
Selenium (µg/L)	1.50	7.60	0.34
Silver (µg/L)	1.25	2.00	0.09
Zinc (µg/L)	8.13	157.00	0.66

Table 3-2. Confusion matrices for predictions of biological condition.

		O/E				MMI			
		Predicted				Predicted			
		Reference	Degraded	Class Error		Reference	Degraded	Class Error	
Observed	Reference	339	29	8%	Observed	Reference	257	62	19%
	Degraded	66	35	67%		Degraded	81	69	54%

Table 3-3. Summary of RF models predicting biologically important stressors. RF % Var = percent of variation in each stressor that was explained by the model. Predictors are listed from left to right in order of importance. Signs in parentheses indicate the general direction of the stressor in response to a predictor. No sign indicates a complex or ambiguous relationship.

Stressor	RF % Var	Predictors
EC	96	Predicted EC (+), % watershed agriculture (+), mine denisty (+), % watershed urban (+)
Temperature	52	Predicted stream temperature (+), maximum dam volume (-), % watershed Ag (+)
Total N	47	% Watershed urban (+),% watershed Ag (+), mine denisty (+), predicted total N (+)
Total P	50	% watershed agriculture (+), % watershed urban (+), predicted total P (+)
Arsenic	32	watershed mean annual maximum precipitation (-), site elevation (-), watershed mean annual minimum temperature (+)
Copper	80	% watershed urban (+), watershed mean annual maximum wet days (-), site elevation (-), watershed mean hydrologic stability (-), mine denisty (+)
Zinc	62	% watershed urban (+), mine denisty (+), site elevation (-)

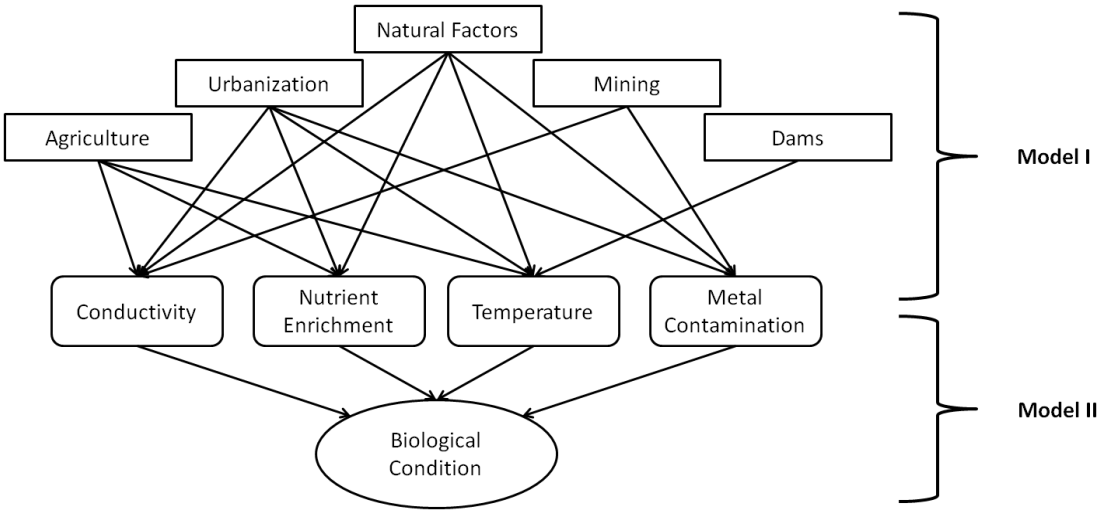


Fig. 3-1. Conceptual model identifying possible linkages between land uses, instream stressors, and biological condition.

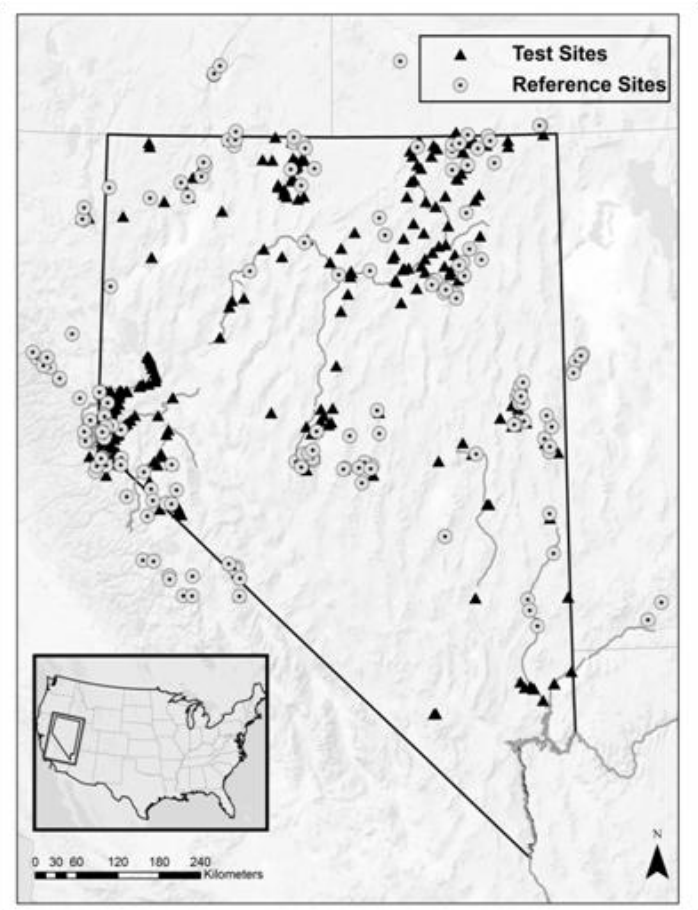


Fig. 3-2. Benthic invertebrate sampling locations. Some sites in near proximity to NV were included for index development and evaluations.

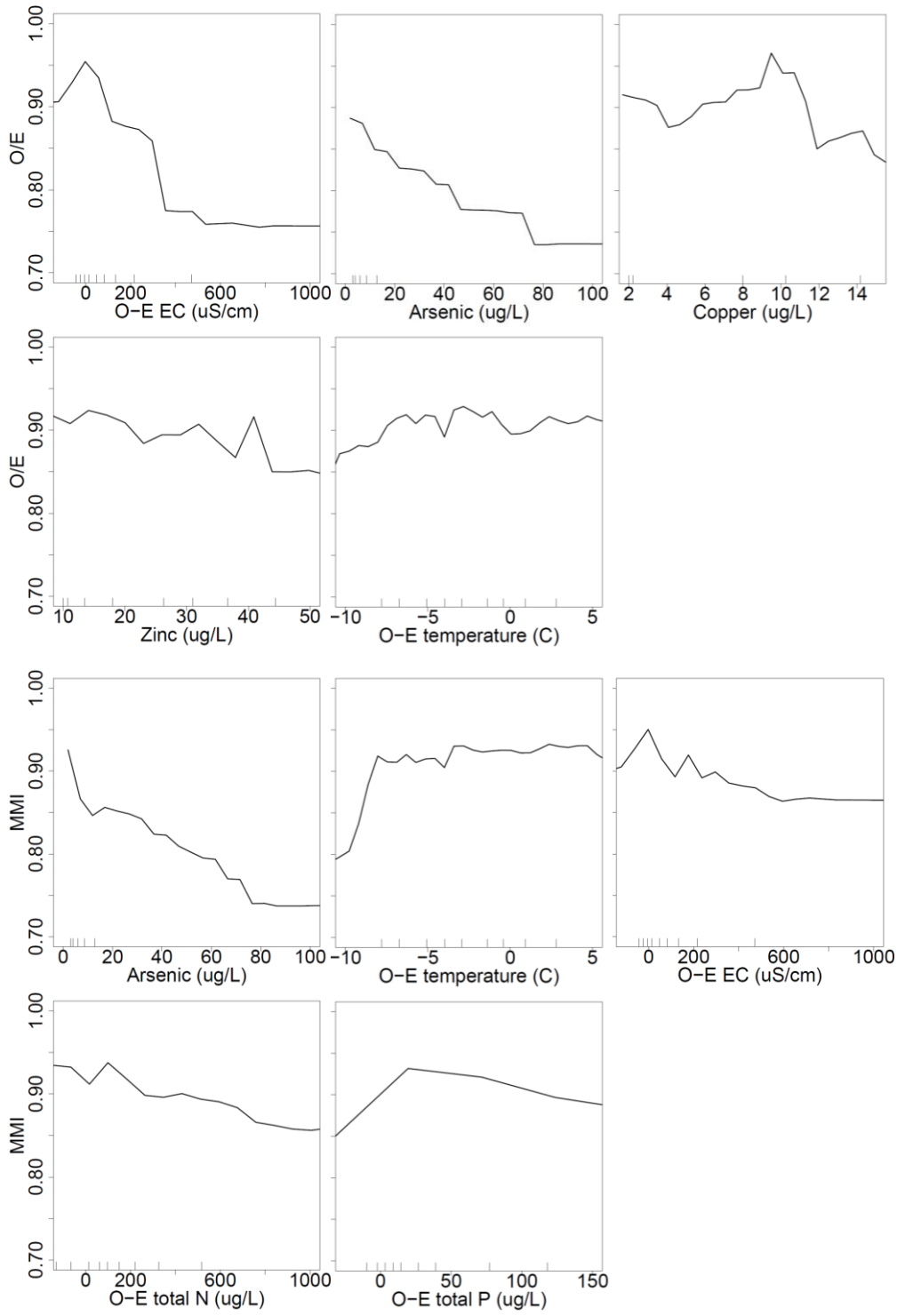


Fig. 3-3. Partial dependence plots of O/E_{taxa} (left) and MMI (right) all stressors used in biota-stressor models.

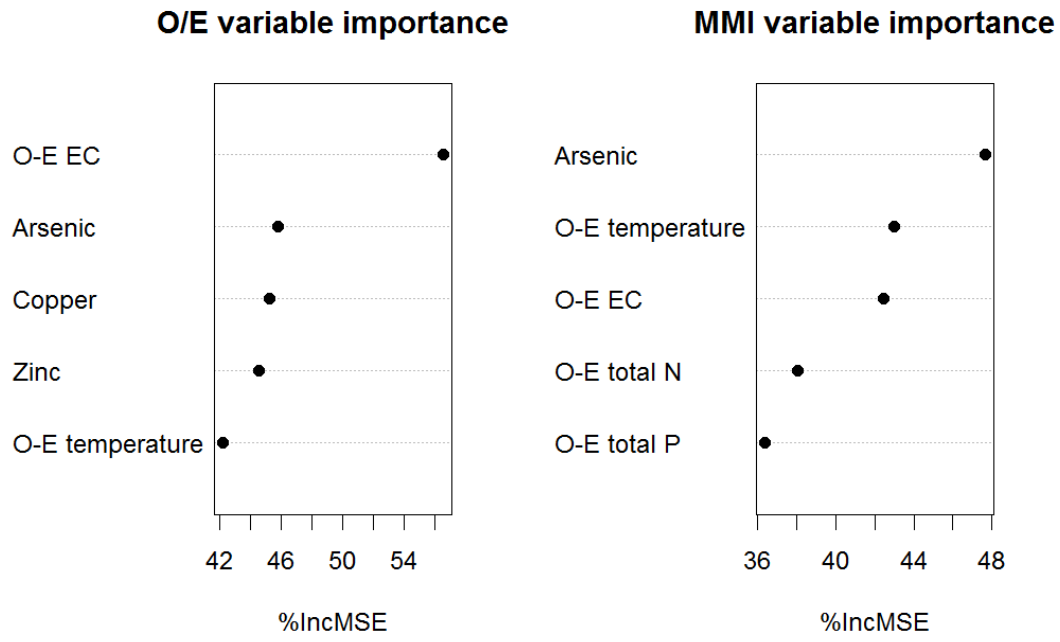


Fig. 3-4. Variable importance plots from biota-stressor RF models.

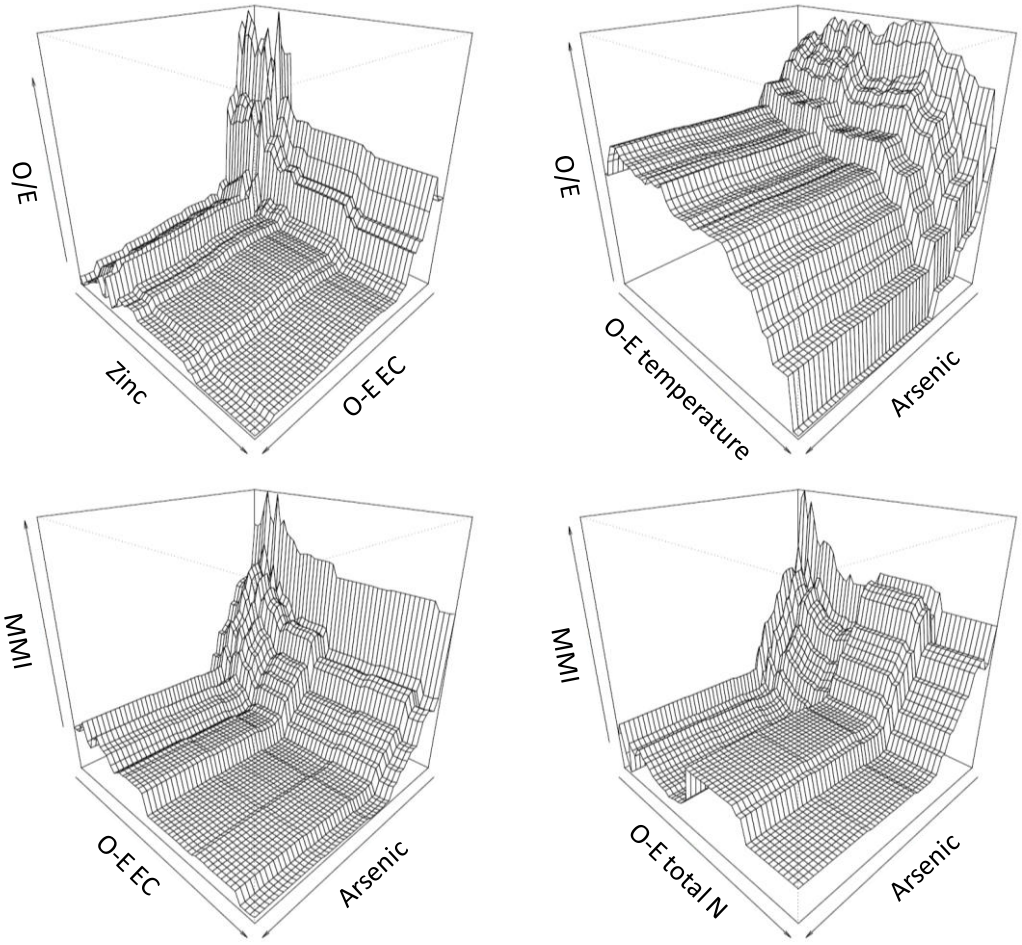


Fig. 3-5. Selected bivariate partial dependence plots showing the joint effects of pairs of stressors on O/E_{taxa} (top) and MMI (bottom).

CHAPTER 4

SUMMARY AND CONCLUSIONS

Managers of stream ecosystems are charged with the tasks of assessing, protecting, and, where necessary, restoring the biological integrity of streams. To do this, they must not only have tools to detect degradation, but also to identify likely causes of degradation. My thesis work focused on developing tools that can be used to measure biological condition, detect biological impairment, infer likely causes of degradation in Nevada streams, and identify potential sources of stressors. These tools should help managers to protect freshwater resources by improving their ability to identify biological degradation and its likely causes.

Bioassessments are commonly used to measure biological condition and detect degradation of streams, and are generally based on one of two types of biological indices: multimetric indices (MMI) and observed to expected (O/E) taxa ratios (Cao and Hawkins 2011). However, the performance of the biological indices we develop influences our ability to detect biological impairment, and although performance of indices can vary widely, the causes of this variability are of often unclear. We developed both multimetric and O/E indices to assess the biological condition of Nevada streams and determined whether index performance was associated with site isolation, beta diversity, and sample evenness. Sample evenness was positively associated with both estimates of richness and index scores. Post-hoc adjustments of index scores for sample evenness resulted in improved index precision, but also altered inferences of biological condition for several sites. Improved sampling may minimize the effects of sample evenness on index precision, but the biogeographic factors that influence the spatial distribution of biota in a region may still impose limits on the ability to predict assemblage composition and thus the performance of bioassessment indices.

Bioassessments may be particularly difficult to develop for generally arid, but highly heterogeneous regions like Nevada. The O/E indices we developed for Nevada were relatively imprecise compared with those developed for other regions. This relative imprecision may be the consequence of low assemblage predictability associated with spatial isolation of aquatic habitats in arid regions. In contrast, the Nevada MMI, which was based on a modeling approach to account for the variability of assemblage properties along environmental gradients, was much more precise than most previously developed MMIs.

Although there are many stressors known to cause stream degradation, identifying the stressors responsible for degradation at specific sites is problematic. Inferences regarding the specific stressors that are causing degradation can be challenging because stressors often co-occur and can have interactive effects (Allan 2004, Townsend et al. 2008). In addition these types of cause and effect relationships can rarely be established in single studies (Norris et al. 2012). Because of these challenges, determining causes of degradation often requires a weight of evidence approach (e.g. Suter et al. 2010, Allan et al. 2011, Norris et al. 2012). In these approaches, causation can be inferred if there is a plausible causal mechanism and a consistent and strong association between the hypothesized cause and effect (Norris et al. 2012). In our study, conductivity and metal contamination were the stressors most strongly associated with biological degradation. Chronic metal contamination is known to cause several harmful effects on aquatic invertebrates (e.g. Thorp et al. 1979, Wicklum and Davies 1996, Richardson and Kiffney 2000, Clements 2004, Carlisle and Clements 2005). These established mechanisms combined with our finding of a strong association between metal contamination and degradation provides support for the inference that metal contamination is a stressor of concern in Nevada streams. The effects of conductivity on stream biota are less well established,

but conductivity was the one of the most important predictor of biological degradation in Nevada streams. However, the levels of conductivity at which we observed degradation are well below levels associated with direct toxicity of freshwater invertebrates. Our results suggest that degradation associated with elevated conductivity does not result from toxicity, but instead from changes in chemical niche space that result in changes in assemblage composition.

Because important stressors like conductivity and metal contamination tended to be associated with land uses, the protection or restoration of biological integrity will likely require management techniques that mitigate the delivery of these stressors from sources to streams. The stressors that we identified may also be important in other landscapes. For example, elevated conductivity in streams resulting from mountaintop mining in Appalachian streams (e.g. Pond 2010, U.S. EPA 2011) and metal contamination in regions with widespread mining (e.g. Clements et al. 2000, Pollard and Yuan 2006) have both been linked to changes in invertebrate assemblages and overall stream degradation. Finally, although conductivity can be linked to stream degradation, it is not currently regulated by water quality standards, and managers should consider ways to establish conductivity criteria for streams.

Literature Cited

ALLAN, J. D. 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. Annual Review of Ecology Evolution and Systematics 35:257-284.

ALLAN, J.D., L.L. YUAN, P. BLACK, T. STOCKTON, P.E. DAVIES, R.H. MAGIEROWSKI, AND S.M. READ. 2011. Investigating the relationships between environmental stressors and the stream condition using Bayesian belief networks. Freshwater Biology doi: 10.1111/j.1365-2427.2011.02683.x.

CAO, Y., AND C. P. HAWKINS. 2011. The comparability of bioassessments: a review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680-701.

CARLISLE, D. M., AND W. H. CLEMENTS. 2005. Leaf litter breakdown, microbial respiration and shredder production in metal-polluted streams. Freshwater Biology 50:380-390.

CLEMENTS, W. H. 2004. Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. Ecological Applications 14:954-967.

CLEMENTS, W. H., D. M. CARLISLE, J. M. LAZORCHAK, AND P. C. JOHNSON. 2000. Heavy metals structure benthic communities in Colorado mountain streams. Ecological Applications 10:626-638.

NORRIS, R.H., J.A. WEBB, S.J. NICHOLS, M.J. STEWARDSON, AND E.T. HARRISON. 2012. Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshwater Science 31:5-21.

POLLARD, A. I., AND L. YUAN. 2006. Community response patterns: evaluating benthic invertebrate composition in metal-polluted streams. Ecological Applications 16:645-655.

POND, G. J. 2010. Patterns of Ephemeroptera taxa loss in Appalachian headwater streams (Kentucky, USA). Hydrobiologia 641:185-201.

RICHARDSON, J. S., AND P. M. KIFFNEY. 2000. Responses of a macroinvertebrate community from a pristine, southern British Columbia, Canada, stream to metals in experimental mesocosms. Environmental Toxicology and Chemistry 19:736-743.

SUTER, G. W., II, S. B. NORTON, AND S. M. CORMIER. 2010. The science and philosophy of a method for assessing environmental causes. Human and Ecological Risk Assessment 16:19-34.

THORP, J. H., J. P. GIESY, AND S. A. WINERITER. 1979. Effects of Chronic Cadmium Exposure on Crayfish Survival, Growth, and Tolerance to Elevated Temperatures. Archives of Environmental Contamination and Toxicology 8:449-456.

TOWNSEND, C. R., S. S. UHLMANN, AND C. D. MATTHAEI. 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45:1810-1819.

U.S. EPA (ENVIRONMENTAL PROTECTION AGENCY). 2011. A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams. Office of Research and Development, National Center for Environmental Assessment, Washington, DC. EPA/600/R-10/023.

WICKLUM, D., AND R. W. DAVIES. 1996. The effects of chronic cadmium stress on energy acquisition and allocation in a freshwater benthic invertebrate predator. Aquatic Toxicology 35:237-252.

APPENDICES

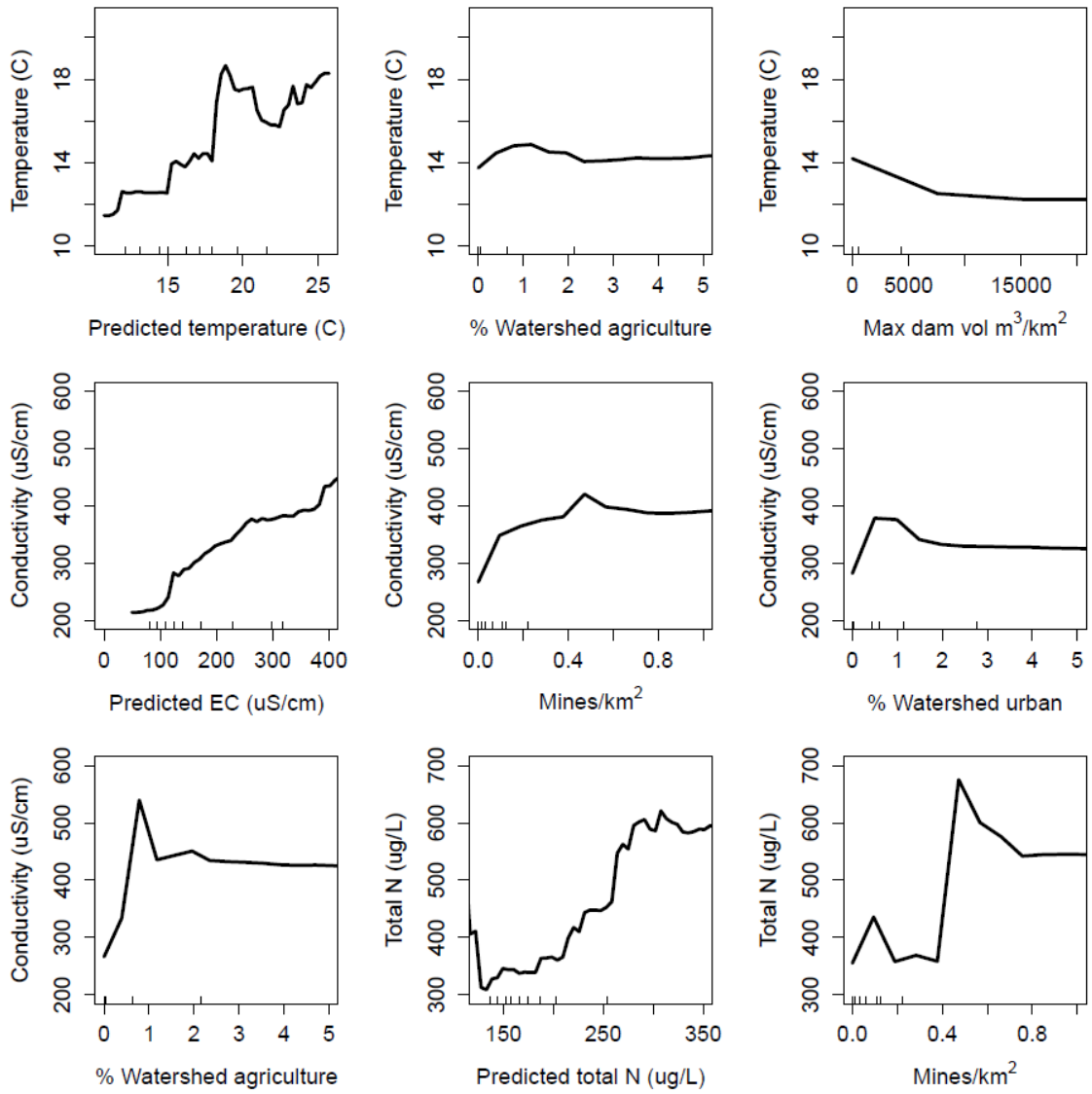
Appendix A. Natural and anthropogenic variables used as potential predictors of stressors.

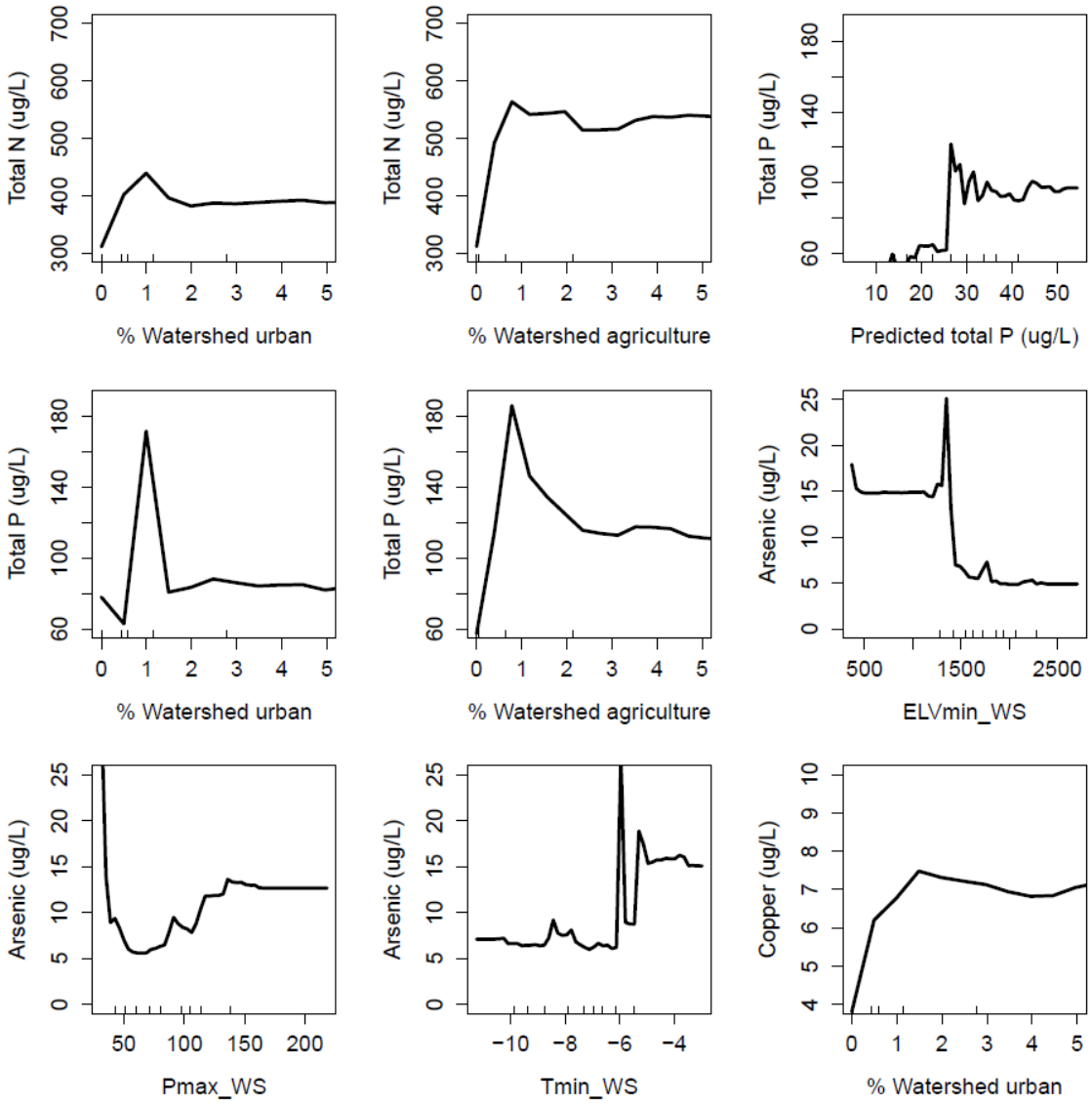
Variable	Description
AG_WS	Percent of watershed classified as agriculture (Homer et al. 2007).
BFI_WS	Mean of all base flow index pixel values within the watershed. Estimates the percent of stream flow that is composed of ground water relative to event flow. Calculated from USGS generated 1-Km resolution grid of base flows derived by interpolating calculated base flows at 19,000 USGS stream flow gauging stations distributed across the conterminous USA (Wolock, 2003)
CaO_Mean	Mean of all cells within the watershed, where cells represent the percent of the underlying bedrock composed of calcium oxide (CaO). Percentages are the average percent CaO for all lithologies within a cell, weighted by lithology prevalence. Lithologies and their prevalences were derived from the USGS Preliminary Integrated Geologic Map of the United States.
DOM_GEOL	Geology type with largest percent coverage within the watershed derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States.
ELVcv_PT	Coefficient of variation of elevations within a radius of 5 digital elevation model cells (30 x 30 meter resolution) of the sample site.
ELVmax_WS	Maximum watershed elevation in meters
ELVmean_WS	Mean watershed elevation in meters
ELVmin_WS	Minimum watershed elevation in meters
HYDR_PT	GIS raster calculated as ($MINx_i$) / ($MAXx_i$), where x_i = mean monthly discharge for month i for the period of record and $x_i \geq 12$ months of record. Values were calculated for each of 9,941 USGS gauging stations in the western USA and values for unmeasured locations were interpolated using inverse-distance-squared weighting of the 12 closest gauging stations within 100 kilometers. Each interpolated value represents a 4 x 4 kilometer cell.
HYDR_WS	Mean of all HYDR_PT values within the watershed
NEAR_AG	Percent of watershed within 5 km of sample location classified as agriculture (Homer et al. 2007).
NEAR_URB	Percent of watershed within 5 km of sample location classified as urban (Homer et al. 2007).
Max_DamVol	The volume (km^3) of the largest dam within the watershed (USACE 2006). Standardized by dividing by watershed area.
MgO_Mean	Mean of all cells within the watershed, where cells represent the percent of the underlying bedrock composed of magnesium oxide (MgO). Percentages are the average percent MgO for all lithologies within a cell, weighted by lithology prevalence. Lithologies and their prevalences were derived from the USGS Preliminary Integrated Geologic Map of the United States.
MINEperSQKM	Watershed mine density calculated as the number of mines divided watershed area (USGS mineral resources data system, http://mrdata.usgs.gov).

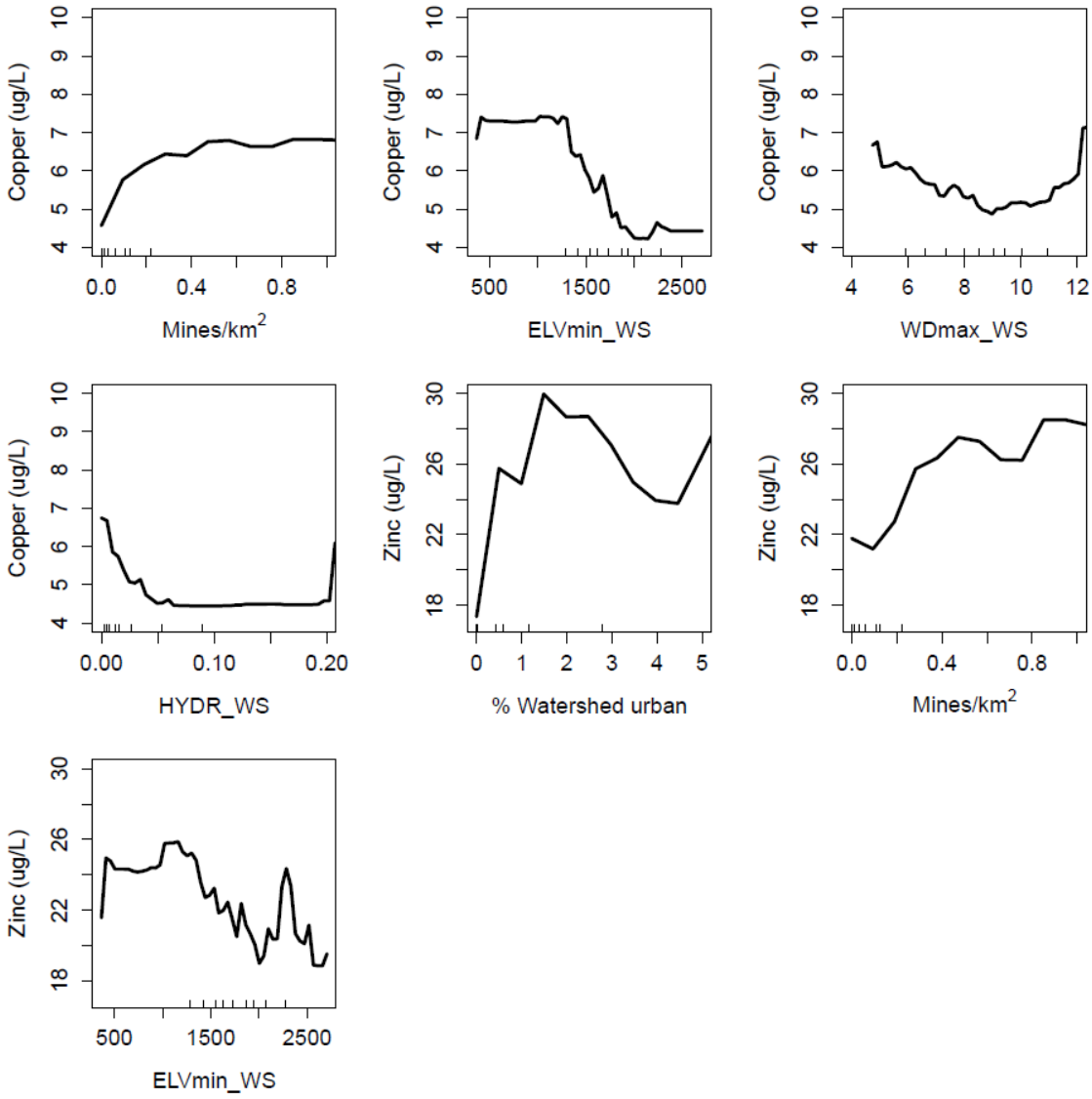
Pmax_PT	GIS raster calculated as S(MAXx_i) / 30 at the sampling point, where xi = the modeled total precipitation (mm) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).
Pmax_WS	Mean of all Pmax_PT values within the watershed
Pmin_PT	GIS raster calculated as SMINx_i / 30 at the sampling point, where $x_{i\,=}$ the modeled total precipitation (mm) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).
Pmin_WS	Mean of all Pmin_PT values within the watershed
PrdCond	Expected conductivity at sampling point. (Olson and Hawkins 2012)
PrdTemp	Expected summer stream temperature (˚C) (Hill, USU, unpublished).
RHmean_PT	GIS raster calculated as S(SUMx_i / 12) / 30 at the sampling point, where $x_{i\,=}$ the modeled mean relative humidity (%) for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 kilometer cell (http://www.prism.oregonstate.edu).
RHmean_WS	Mean of all Rhmean_PT values within the watershed
S_Mean	Mean of all cells within the watershed, where cells represent the percent of the underlying bedrock composed of sulfur (S). Percentages are the average percent S for all lithologies within a cell, weighted by lithology prevalence. Lithologies and their prevalences were derived from the USGS Preliminary Integrated Geologic Map of the United States.
Slope_WS	Watershed slope measured as the (ELVmax_WS-ELVmin_WS)/Maximum flow length. Calculated from statistics produced by the multi-watershed delineation tool (Chinnayakanahalli et al. 2006).
Sum_DamVol	The total volume (km^3) of all dams within the watershed (USACE 2006). Standardized by dividing by watershed area.
Tmax_PT	GIS raster calculated as SMAXx_i / 30 at the sampling point, where x_i = the modeled monthly average maximum air temperature (°C) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell. Note that these values are modified from the PRISM annual maximum air temperature grid available at: http://www.prism.oregonstate.edu, that are calculated as S(SUMx_i / 12) / 30, where x_i = the modeled monthly average maximum air temperature (°C) for month i (1-12).
Tmax_WS	Mean of all Tmax_PT values within the watershed
Tmean_PT	GIS raster calculated as S(SUMx_i / 12) / 30 at the sampling point, where x_i = the modeled mean air temperature (°C) for month i (1-12). The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).
Tmean_WS	Mean of all Tmean_PT values within the watershed

Tmin_PT	GIS raster calculated as SMINx_i / 30 at the sampling point, where x_i = the modeled monthly average minimum air temperature (°C) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell. Note that these values are modified from the PRISM annual maximum air temperature grid available at: http://www.prism.oregonstate.edu, that are calculated as S(SUMx_i / 12) / 30, where x_i = the modeled monthly average minimum air temperature (°C) for month i (1-12).
Tmin_WS	Mean of all Tmin_PT values within the watershed
UCS_Mean	Mean of all cells within the watershed, where cells represent the average of uniaxial compressive strength (UCS, MPa) of the underlying bedrock. Cell values are the average UCS for all lithologies within that cell, weighted by lithology prevalence. Lithologies and their prevalences were derived from the USGS Preliminary Integrated Geologic Map of the United States.
URB_WS	Percent of watershed classified as urban (Homer et al. 2007).
WSA	Watershed area in square kilometers

Appendix B. Partial dependence plots of all stressors on land uses and natural characteristics.




Total N (ug/L)
% Watershed urban
Total N (ug/L)
% Watershed agriculture
Total P (ug/L)
Predicted total P (ug/L)
Total P (ug/L)
% Watershed urban
Total P (ug/L)
% Watershed agriculture
Arsenic (ug/L)
ELVmin_WS
Arsenic (ug/L)
Pmax_WS
Arsenic (ug/L)
Tmin_WS
Copper (ug/L)
% Watershed urban


Copper (ug/L)
4 5 6 7 8 9 10
0.0 0.4 0.8
Mines/km2
Copper (ug/L)
4 5 6 7 8 9 10
500 1500 2500
ELVmin_WS
Copper (ug/L)
4 5 6 7 8 9 10
4 6 8 10 12
WDmax_WS
Copper (ug/L)
4 5 6 7 8 9 10
0.00 0.10 0.20
HYDR_WS
Zinc (ug/L)
18 22 26 30
0 1 2 3 4 5
% Watershed urban
Zinc (ug/L)
18 22 26 30
0.0 0.4 0.8
Mines/km2
Zinc (ug/L)
18 22 26 30
500 1500 2500
ELVmin_WS